

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INDUSTRIALES Y DE TELECOMUNICACIÓN

UNIVERSIDAD DE CANTABRIA



Trabajo Fin de Grado

**ASIGNACIÓN DE FUNCIONES DE RED EN
REDES 5G**

**(Network function assignment in 5G
networks)**

Para acceder al Título de

***Graduado en
Ingeniería de Tecnologías de Telecomunicación***

Autor: David Ruiz Soto

09 - 2021



E.T.S. DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACION

GRADUADO EN INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN

CALIFICACIÓN DEL TRABAJO FIN DE GRADO

Realizado por: David Ruiz Soto

Director del TFG: Luis Francisco Díez Fernández, Ramón Agüero Calvo

Título: “ Asignación de funciones de red en redes 5G”

Title: “Network function assignment in 5G networks “

Presentado a examen el día: 28/09/2021

para acceder al Título de

GRADUADO EN INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN

Composición del Tribunal:

Presidente: Pereda Fernández, José Antonio

Secretario: Díez Fernández, Luis Francisco

Vocal: Sanz Gil, Roberto

Este Tribunal ha resuelto otorgar la calificación de:

Fdo.: El Presidente

Fdo.: El Secretario

Fdo.: El Vocal

Fdo.: El Director del TFG
(sólo si es distinto del Secretario)

Vº Bº del Subdirector

Trabajo Fin de Grado N°
(a asignar por Secretaría)

Índice general

1. Introducción	7
1.1. Motivación	7
1.2. Objetivos	8
1.2.1. Configurar la herramienta de análisis	8
1.2.2. Creación de los algoritmos	9
1.2.3. Extraer datos de la herramienta	9
1.2.4. Analizar los datos obtenidos	9
1.3. Estructura de la memoria	9
1.3.1. Conceptos previos y contexto del trabajo	9
1.3.2. Presentación del trabajo y sus distintos elementos	10
1.3.3. Funcionamiento del programa	10
1.3.4. Análisis de resultados y conclusiones finales	10
2. Conceptos previos	11
2.1. Telefonía móvil	11
2.2. Estructura de la red	13
2.2.1. BackHaul de la red móvil	14
2.2.2. FrontHaul de la red móvil	14
2.3. Evolución de la tecnología	15
2.3.1. Primera Generación	15
2.3.2. Segunda generación	16
2.3.3. Generación 2.5	19
2.3.4. Tercera Generación	20
2.3.5. Cuarta Generación	23

2.4. Redes 5G	24
3. Problema a resolver	27
3.1. Presentación escenario	28
4. Implementación	31
4.1. Herramienta	32
4.2. Escenario	32
4.3. Creación de la red	33
4.4. Análisis de la red	34
4.5. Final de la herramienta	34
5. Explicación de los algoritmos	35
5.1. Algoritmo Paths	35
5.1.1. Funcionamiento del algoritmo	36
5.2. Algoritmo Arbol Soluciones	40
5.2.1. Funcionamiento del algoritmo	41
5.3. Algoritmo Escoger Solución	50
5.3.1. Funcionamiento del algoritmo	51
6. Datos y Resultados	57
6.1. Datos con los que se ejecuta el programa	57
6.1.1. Tecnologías de enlace	57
6.1.2. Niveles de split	58
6.1.3. Topologías de red	58
6.2. Resultados de la ejecución	61
6.2.1. Delay	62
6.2.2. Nivel de split	63
6.3. Análisis de los Resultados	65
6.3.1. Delay	66
6.3.2. Nivel de Split	71
6.3.3. Criterios de selección de solución	76
7. Conclusiones y líneas futuras	77

A. Terminos utilizados	81
A.1. Herramienta de análisis	81
A.2. Programa	81
A.3. Escenario	82
A.4. Creación de la red	82
A.5. Asignación de los nodos	82
A.6. Algoritmos	82
A.7. Criterios	83
A.8. Árbol de soluciones	83
A.8.1. Vector	83
A.8.2. Mapa	83
A.8.3. Estructura	84
A.8.4. Solución	84
A.8.5. Decisión	84
A.8.6. Explicación del árbol de soluciones	84
A.9. Métricas	85
A.10.Ejecución	85
A.11.Acción	86
A.12.Topología	86
A.13.Red	86
A.14.Fotografía	87
A.15.Nodos	87
A.15.1.RUs	87
A.15.2.CUs	87
A.15.3.INTs	87
A.16.Funcionamiento	87
A.17.Enlaces	88
A.18.Caminos	88
A.19.Métricas de los enlaces y caminos	88
A.19.1.Delay	89
A.19.2.Capacidad	89

A.20.Nivel de centralización	89
--	----

Capítulo 1

Introducción

1.1. Motivación

El contexto en el que se desarrolla este trabajo, es la reciente introducción de las redes de telefonía móvil de quinta generación o 5g. Este nuevo tipo de redes, que aún se encuentran en un estado de instalación inicial, ofrecen una mejora en la capacidad de transmisión de datos y en la latencia en el envío de los mismos, y con ello una mejora en los servicios que los proveedores de datos móviles pueden ofrecer a sus usuarios; así como un aumento en la cantidad de usuarios que pueden encontrarse conectarse simultáneamente a la red. Además de estas mejoras, también podemos observar varias diferencias de funcionamiento con respecto a sus predecesoras.

Uno de los cambios más importantes, y sobre el que trata este trabajo, es la asignación de las funciones de red de las distintas estaciones de radio, que como se puede entender, son una serie de funciones que cada estación de radio necesita que se ejecuten, para que se pueda realizar esa comunicación y envío de datos. Estas funciones pueden ir desde la encriptación y procesamiento de los datos a enviar, hasta el encaminamiento de los mismos, pasando por la autenticación de los usuarios, entre otras cosas.

En las redes de cuarta generación, o 4G, estas funciones se ejecutaban de forma completamente distribuida, de forma que cada una de las estaciones de radio de una red se encargaba de ejecutar sus propias funciones. Y en general, en todas las generaciones anteriores a la quinta, todas las funciones se ejecutaban en un mismo lugar, ya fuera completamente distribuidas en cada estación de radio, o centralizadas en un controlador de red que se encargase de ejecutar las funciones de un grupo de estaciones de radio.

En cambio, con la aparición de la quinta generación, que trae asociado el uso de la nube, no es necesario que exista un controlador de red que gestione varias estaciones, y por tanto, se abre la posibilidad a centralizar una parte variable de las funciones de red, siempre que la conexión de la estación de radio con la nube posea la capacidad necesaria para enviar los datos requeridos para realizar estas funciones.

Con esto, en las redes 5g, las funciones de red ya no se realizan de forma completamente distribuida ni centralizada, sino que en cada caso, y en cada estación de radio, se pueden centralizar una mayor o menor cantidad de funciones, dependiendo de las características y el estado de la red, de forma que en una misma red, podemos encontrarnos con una estación de radio que ejecuta casi todas las funciones de red de forma distribuida, mientras otra centraliza un mayor número de funciones, siempre dependiendo del estado de la red, o de lo que sea más conveniente en cada caso.

A la cantidad de funciones que se ejecutan en la nube de cada estación de radio, le llamaremos nivel de centralización de esa estación de radio. Y en esta situación es interesante tener una herramienta que permita determinar cuál es el nivel de centralización óptimo para las distintas estaciones de radio de una red, en unas condiciones dadas, y según los criterios que consideremos más relevantes.

El objetivo de este trabajo es, por tanto, crear una herramienta de análisis que permita encontrar la mejor asignación de este nivel de centralización para las distintas estaciones de radio de una red dada, dependiendo de las exigencias de las distintas funciones para ser centralizadas, y de las características de los enlaces que conecten las estaciones de radio con la nube.

Concretamente, el algoritmo desarrollado, realiza una búsqueda completa entre todas las combinaciones posibles, comparando todas las soluciones hasta encontrar la mejor en cada caso según el criterio o criterios seleccionados. Por este motivo, y al crecer el número de combinaciones de forma exponencial, a poco que crezcan las redes, el tiempo necesario para realizar esta búsqueda se vuelve demasiado alto. Por este motivo, el objetivo este trabajo no es crear una herramienta que pueda funcionar a tiempo real, sino un marco de referencia con el que comparar otras herramientas que sí lo hagan ofreciendo soluciones aproximadas, permitiendo comparar la utilidad y la precisión de estas con respecto a la solución exacta.

Además, al tener una herramienta que funciona de forma exacta, también es interesante poder analizar la forma en la que los cambios en las características en las redes afectan a los niveles de centralización que se pueden asignar a las estaciones de radio de las mismas, de forma que además de una parte de programación, el trabajo también incluye una parte de análisis de datos.

1.2. Objetivos

Los objetivos del trabajo se encuentran directamente relacionados con la motivación del mismo, puesto que buscamos precisamente resolver las cuestiones que la aparición de las redes 5G nos hacen plantearnos. Con esto se puede entender que el objetivo principal de este trabajo es crear una herramienta que permita asignar la centralización de las funciones de red según los criterios seleccionados, y analizar los resultados de la ejecución de esta herramienta sobre distintas redes. Aun así, en esta sección explicaré cada uno de los objetivos de forma pormenorizada.

1.2.1. Configurar la herramienta de análisis

El objetivo principal de este trabajo, como ya se ha explicado, es crear una herramienta de análisis que permita encontrar la mejor combinación de niveles de centralización para las distintas estaciones de radio en una red dada, pues es precisamente esta nueva forma de asignar las funciones la novedad de las redes 5G en el la que se centra este trabajo.

Además, para que el programa se pueda ejecutar sobre una red, es necesario crear un sistema que permita introducir estas redes en la herramienta, por lo que esta también será una parte del desarrollo. Y además, con el objetivo de poder extraer resultados a una escala que permita que posteriormente puedan ser analizados, esta parte de la herramienta incluye la opción de crear redes distintas sobre una misma topología de red (que tendría que ser introducida antes de ejecutar el programa). Para esto asigna tanto los nodos como las características de los enlaces de forma aleatoria, pero permitiendo que esta aleatoriedad sea controlada por el usuario mediante un fichero de configuración en el que indique los valores entre los que se deben mover los distintos parámetros de la red generada.

1.2.2. Creación de los algoritmos

Esta se puede considerar una parte de la configuración de la herramienta de análisis, pero es importante remarcar que además de un sistema que permita introducir y procesar los datos, es necesario desarrollar los algoritmos que permitan que la herramienta cumpla sus funciones. Estos algoritmos se encargan tanto de generar todas las posibles combinaciones, como de compararlas según una serie de criterios, e incluye la posibilidad de añadir criterios distintos para seleccionar la mejor solución, según las necesidades concretas de cada caso.

1.2.3. Extraer datos de la herramienta

El tercer objetivo está precisamente ligado con la posibilidad que nos da la herramienta de generar varias redes sobre una topología, pues no es otro que extraer datos de las soluciones obtenidas, tras ejecutar la herramienta sobre distintas redes.

De esta forma el objetivo es, además de generar unos resultados que puedan servir como referencia a la hora de analizar otras herramientas con funcionamiento a tiempo real, poder comparar los niveles de centralización que se asignan a distintas redes según sus características, o según el criterio empleado para escoger cuál es el nivel de centralización en cada caso. Y para hacerlo se genera un cierto número de redes, cambiando la asignación de sus nodos o su tecnología de enlace, y se extraen una serie de métricas de las soluciones obtenidas al ejecutar la herramienta de análisis sobre cada una de ellas.

1.2.4. Analizar los datos obtenidos

El último objetivo es, por tanto, analizar los datos de esas métricas obtenidas, en este caso representándolos gráficamente, de forma que podamos analizar y explicar de qué forma cambian los resultados, según el tipo de red, su tamaño, tecnología de enlace, o el criterio de selección de la solución en cada caso.

1.3. Estructura de la memoria

El objetivo de esta memoria es explicar tanto el contexto en el que se han desarrollado la herramienta y el trabajo, como el funcionamiento de esta, y el análisis de los datos que se han extraído tras ejecutar el programa sobre un número razonablemente alto de redes. Para poder explicar cada uno de estos puntos de forma más clara, la memoria se divide en las siguientes partes principales:

1.3.1. Conceptos previos y contexto del trabajo

La primera parte de la memoria se encarga de explicar tanto el contexto previo necesario para entender el sentido del trabajo, como la situación actual en la que tiene sentido esta herramienta de análisis. Esto es, a grandes rasgos, una explicación de la evolución de las telefonía móvil a lo largo de sus distintas generaciones antes de llegar a las redes 5G, así como las funciones de red que era necesario ejecutar en las distintas generaciones y la forma en la que se repartían. Y por el otro lado, el estado actual de la tecnología con la aparición de estas redes 5G, y la forma en la que esto motiva la realización de este trabajo.

1.3.2. Presentación del trabajo y sus distintos elementos

La segunda parte se encarga de presentar las partes principales de la herramienta, incluyendo el problema que busca solucionar, el tipo de redes sobre las que vamos a ejecutar el programa, y el funcionamiento general del mismo.

1.3.3. Funcionamiento del programa

En la siguiente parte se explica el funcionamiento detallado de la herramienta de análisis y de las distintas acciones de las que está compuesta. En esta parte se explica como gestiona la herramienta las distintas topologías sobre las que se va a ejecutar el programa, como controla el orden en el que se ejecutan las acciones, como se generan las redes de las que vamos a extraer las soluciones, y cómo funcionan también cada uno de los algoritmos que en conjunto permiten que el programa funcione y devuelva los resultados esperados.

1.3.4. Análisis de resultados y conclusiones finales

Finalmente, la última parte de la memoria muestra los resultados extraídos de la ejecución de la herramienta, y a qué conclusiones podemos llegar a partir de ellos. Para hacerlo, se explicarán las distintas topologías y condiciones de las redes sobre las que se han realizado los análisis, y se mostrarán y estudiarán las gráficas con las métricas extraídas de las distintas soluciones. Una vez hecho esto, se sacarán conclusiones de los resultados obtenidos, y se plantearán las líneas de futuro que podemos encontrar a partir de este trabajo.

Capítulo 2

Conceptos previos

Antes de explicar el programa, su funcionamiento, y los datos que hemos extraído del mismo, el paso previo es entender el contexto de la tecnología en la que se desarrolla, que son las redes de quinta generación o 5G, así como las ventajas de esta, y la nueva forma de organización de las redes que se introduce con esta generación en comparación con las anteriores generaciones de telefonía.

Para poder realizar esta explicación de la mejor forma, en este capítulo se hablará de la evolución de la tecnología móvil desde sus inicios, explicando el funcionamiento general de las distintas generaciones, así como los cambios, tanto en funcionamiento y tecnologías empleadas, como en servicios y capacidades que se van produciendo entre ellas, hasta llegar a las redes de quinta generación o 5G.

Adicionalmente, y antes de explicar cada una de las generaciones de telefonía móvil por separado, se explicará de forma bastante simplificada el funcionamiento general de las redes de telefonía móvil, pues esto facilitará explicar el funcionamiento de las distintas generaciones de la misma, así como a dar un contexto sobre qué quieren decir los distintos cambios, o cuales son las distintas partes de una red de telefonía móvil 5G, y sobre qué parte de estas funciona el programa.

Por tanto, este capítulo se dividirá en tres partes, primero una explicación general de la telefonía móvil, a continuación un resumen de cada una de las generaciones previas a las redes 5G, que sirva como contexto para esta, y por último, una explicación más extensa de estas redes 5G, pues son sobre las que tiene sentido este programa, y sobre las que se desarrolla el trabajo en general.

2.1. Telefonía móvil

La telefonía móvil es un medio de comunicación que funciona de forma inalámbrica, enviando los datos de la comunicación a través de ondas electromagnéticas. Como su propio nombre indica, lo que permite esta telefonía móvil es que, mediante un dispositivo portátil capaz de recibir y enviar ese tipo de señales, y siempre que nos encontremos dentro del rango de cobertura de una antena de radiotelefonía que forme parte de la red, podamos comunicarnos en movimiento desde y con cualquier lugar del mundo.

Estos dispositivos portátiles que permiten que nos conectemos a la red son conocidos como teléfonos móviles, y su funcionalidad no es otra que la de un receptor-emisor de ondas electromagnéticas que contienen la información que se quiere comunicar. Aunque en la práctica, las funciones de los teléfonos móviles no se limitan únicamente al envío y la recepción de datos, sino que han

ido aumentando conforme avanzaban las capacidades de la red.

En sus inicios, el funcionamiento de estos teléfonos consistía en convertir las ondas sonoras de nuestra voz en señales electromagnéticas y enviarlas por el aire a una antena de radio, que a su vez transmitía esta señal a través de la red de telefonía móvil, para que finalmente esta las mandase hasta el teléfono móvil del destinatario, que a su vez las volvía a convertir en ondas sonoras, de forma que el receptor pudiera escuchar lo que habíamos dicho.

En telefonía móvil, una serie de antenas de radio que dan cobertura conjuntamente a una zona concreta, forman una celda o cluster. Estas celdas están organizadas para poder dar cobertura a la totalidad de un territorio o una región determinada, de forma que al movernos de una zona a otra podamos seguir conectados a la red cambiando de una celda a otra, y no haya lugares en los que nos sea imposible comunicarnos. El alcance de estas celdas, y el número de antenas por las que están formadas dependen de cada caso, y también ha ido cambiando con el avance de la tecnología, pero pueden cubrir desde pequeñas regiones hasta grandes extensiones de cientos de kilómetros, dependiendo de la cantidad de dispositivos que queramos conectar, o la potencia de las antenas, entre otros factores.

Generalmente, la forma de estas celdas se considera hexagonal (aunque las antenas tengan un alcance circular), pues de esta forma podemos cubrir la región con el menor número de celdas, al mismo tiempo que permitimos que la distancia entre las antenas sea constante y no nos encontremos con zonas sin cobertura o problemas de mala recepción de la señal. Por otro lado, cada celda utiliza una serie de frecuencias para realizar la comunicación, que no pueden ser empleadas por sus celdas contiguas, de forma que no haya problemas de interferencias en los puntos en los que las señales de dos antenas de celdas distintas se crucen. Es fácil por tanto darse cuenta de que al ser hexagonales, cada celda se encuentra rodeada por otras 6, por lo que cada una utiliza una séptima parte de las frecuencias disponibles para comunicación móvil.

Este ancho de banda (o rango de frecuencias) permitido para cada una de las celdas determina a su vez la cantidad de canales de comunicación, o dispositivos que se pueden conectar simultáneamente dentro de ellas. A este respecto, es necesario tener en cuenta que para evitar interferencias entre dos dispositivos que se encuentran dentro de la misma celda, las señales que envían y reciben deben estar moduladas a frecuencias distintas, siempre dentro del rango permitido para la celda. La cantidad de dispositivos que se pueden conectar, o la capacidad de la red, también depende de la tecnología utilizada en cada caso y la generación en la que nos encontremos, puesto que conforme ha ido avanzando la tecnología, ha sido posible realizar un mayor número de conexiones con mayor capacidad de envío y recepción de datos dentro de una misma celda. Con la tecnología actual, es posible conectar a miles de personas simultáneamente dentro de una misma celda sin que tenga lugar ningún tipo de interferencia entre ellas.

A su vez el grupo de antenas que conforman una celda se encuentra conectada a la red de telefonía, que es la que nos permite conectarnos con cualquier parte del mundo. Por ello, si queremos utilizar nuestro teléfono móvil, debemos de tener cobertura, o lo que es lo mismo, encontrarnos dentro del rango de alcance de una de estas celdas, y que esta tenga capacidad para añadir nuestra conexión a aquellas que ya se encuentran activas en el momento de la conexión.

En las generaciones más avanzadas de telefonía, la comunicación ya no se limita a la conversión directa de forma analógica de las ondas sonoras en ondas electromagnéticas, sino que funciona de forma digital, es decir, que nuestra voz se transforma en una señal digital formada por ceros y unos, que es la que posteriormente se modula a la frecuencia indicada, y se envía por el aire a la red de telefonía para que sea transmitida. Esta digitalización de las señales, nos permite enviar y recibir una variedad mucho más grande de archivos distintos a la voz, y permite que se realicen una serie de operaciones sobre la información enviada para hacer que la comunicación sea más fiable y segura, como puede ser encriptación, confirmación de entrega, etc.

Además, a la hora de realizar la comunicación entre dos dispositivos móviles que se encuentran

conectados a la red de telefonía móvil, es necesario que se ejecuten una serie de funciones de red, como pueden ser el establecimiento de las llamadas, el encaminamiento de las señales a través de la red, etc. Estas funciones de red han ido variando y cambiando conforme avanzaban las distintas generaciones al haber una mayor capacidad tecnológica, y al haberse establecido un mayor número de medidas de seguridad para la comunicación.

Por otro lado, la asignación de los recursos, y el lugar en el que se ejecutaban estas funciones, también ha ido cambiando con las distintas generaciones de telefonía móvil. A grandes rasgos, hasta la tercera generación las funciones de red se centralizaban en un controlador de red, que se encargaba de gestionar una serie de estaciones de radio (normalmente las pertenecientes a una misma celda), mientras que en la cuarta generación, estas funciones de red se ejecutaban en las mismas estaciones de radio. Y por último, como se ha explicado previamente, en la quinta generación, estas funciones de red se pueden ejecutar tanto en las estaciones de radio como de forma centralizada, y el reparto de esa asignación es precisamente el motivo de este trabajo.

2.2. Estructura de la red

En el anterior apartado, se ha explicado el esquema de funcionamiento de la red de telefonía móvil a nivel organización de las estaciones de radio, y más concretamente de las celdas en las que estas se organizan para que los usuarios puedan conectarse con sus dispositivos móviles. Como se explicó en ese apartado, las estaciones de radio se organizan en celdas hexagonales, repartiendo el ancho de banda reservado para la comunicación móvil entre siete, y asignando cada una de esas 7 partes a una celda y a las 6 que la rodean. De esta forma, se evita que los dispositivos de dos celdas adyacentes utilicen las mismas frecuencias, para que no causen interferencias entre ellas.

Dependiendo de la generación de la red y de la tecnología utilizada, cada celda es cubierta por una o varias estaciones de radio, y permite la conexión de un cierto número de usuarios, y la realización de una determinada cantidad de llamadas simultáneamente. Obviamente, el avance de la capacidad de la red, y la evolución de los protocolos de multiplexación permite que el número de canales que se pueden asignar con un mismo ancho de banda sea mucho mayor en las últimas generaciones, y con ello que el número de llamadas que se pueden realizar al mismo tiempo aumente.

Hasta este punto de la sección se ha tratado la estructura de las redes de telefonía móvil de forma global. Pero la parte de estas redes que nos interesa para este trabajo no es la que conecta los teléfonos móviles con las antenas, ni la organización de las celdas, sino la parte de la red que se encarga de conectar dos estaciones de radio distintas entre sí. En el caso de las redes 5G, que son aquellas que queremos estudiar, y sobre las que trabaja la herramienta de simulación que hemos diseñado, esta arquitectura de red se llama C-RAN. Esta arquitectura divide la red en dos partes principales, tal y como se puede ver en la Figura 2.1, llamadas BackHaul y FrontHaul, siendo esta segunda la que nos interesa analizar y la que determina el posible nivel de centralización que podemos emplear en cada estación de radio.

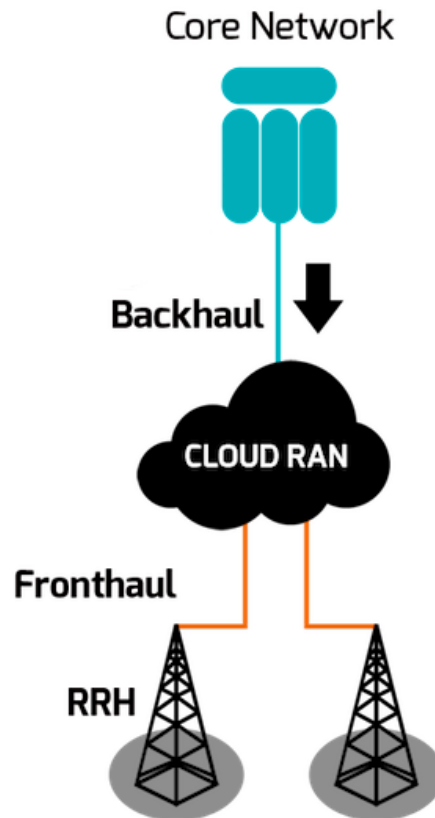


Figura 2.1: Arquitectura C-RAN

2.2.1. BackHaul de la red móvil

El Backhaul de la red de telefonía móvil, como su propio nombre indica, es la parte trasera de la red, o dicho de otra forma, la parte de la red que permite que dos estaciones de radio geográficamente dispersas se conecten entre ellas. Es esta parte de la red la que permite que dos usuarios situados en dos partes completamente distintas del mundo puedan realizar una llamada, o enviarse datos entre ellos.

Su arquitectura y tecnología no son realmente relevantes para la realización de este trabajo, pero a grandes rasgos, la idea es que esta parte de la red posee enlaces de largo alcance y gran capacidad, para permitir el envío de grandes cantidades de datos entre usuarios o servidores que puedan encontrarse muy alejados. En este caso, el Backhaul recurre al estándar Ethernet para interconectar las distintas estaciones de radio, es decir, funciona como si de la red de internet se tratara.

2.2.2. FrontHaul de la red móvil

El Fronthaul es la parte delantera de la red de telefonía móvil. Este trabajo se centra en el funcionamiento y las capacidades de esta, determinando la cantidad de funciones de red que se pueden enviar a la nube, en función de las características y ocupación de los enlaces del fronthaul

de una determinada red. Dicho de otra forma, el fronthaul de la red es el encargado de conectar las estaciones de radio de una determinada ubicación con el backhaul de la red, para luego poder enviar los datos a través de este.

En este trabajo lo que hay que tener en cuenta es que el fronthaul conecta las distintas estaciones de radio con los nodos que dan a la nube, que es donde se pueden ejecutar y centralizar las funciones de red que haya determinado la herramienta de análisis.

2.3. Evolución de la tecnología

Una vez explicado el funcionamiento general de la telefonía móvil, el siguiente apartado consiste en explicar la evolución de esta a lo largo de las distintas generaciones, partiendo desde los primeros teléfonos móviles y el tipo de redes a través de las cuales operaban, hasta llegar a la actualidad.

En esta sección, además de enumerar las distintas funciones y servicios que ofrecían cada una de las generaciones en comparación con sus predecesoras, y de explicar las tecnologías principales en las que se basaban cada una de ellas, se hará especial hincapié en explicar de qué forma y dónde se ejecutaban las funciones de red en cada generación, según las capacidades y las necesidades de cada una de las tecnologías.

2.3.1. Primera Generación

La primera generación, cuyas primeras instancias empezaron a operar en 1979, utilizaba canales analógicos, se basaba en multiplexación FDMA para repartir el ancho de banda entre las distintas llamadas, y únicamente permitía la transmisión de voz. Como se ha explicado en el anterior capítulo, su funcionamiento se limitaba a transformar las ondas sonoras en electromagnéticas, y enviarlas por el aire y a través de la red hasta su destino, es decir, hasta el otro interlocutor con el que estábamos realizando la llamada.

La comunicación era de muy baja calidad, permitiendo envíos de información por debajo de 2,4kbps, sobre todo debido a la poca capacidad de los enlaces que conformaban la red. Además carecía de cualquier tipo de seguridad en la comunicación.

Otro problema en esta red era la comunicación en movimiento, pues al carecer de protocolos de transferencia fiables, era muy probable que al moverse de una celda a otra se perdiera la conexión y fuera necesario volver a realizar la llamada.

FDMA

El acceso múltiple por división de frecuencia, también conocido como FDMA (del inglés Frequency Division Multiple Access) es una técnica de multiplexación usada en varios protocolos de comunicaciones, tanto digitales como analógicos, siendo el protocolo empleado por las primeras redes telefonía móvil.

En FDMA, se divide el espectro disponible, es decir, el ancho de banda que podemos utilizar dentro de una celda de la red de telefonía móvil, en canales, cada uno de los cuales utiliza una frecuencia concreta dentro de ese ancho de banda. Después, se asignan estos canales a los distintos usuarios y a las llamadas que estos quieran realizar, de forma que las señales que cada uno usuarios envía y recibe se encuentran moduladas a una frecuencia concreta, que no es utilizada por ninguna otra comunicación dentro de la celda en la que se encuentra. De esta forma, varios usuarios pueden realizar llamadas dentro de una misma celda simultáneamente sin causarse interferencias entre si.

En algunos sistemas, como es el caso de GSM, el FDMA se complementa con una función de cambio de canal, que permite que una llamada pase a utilizar un canal distinto, según las necesidades de la red en cada momento.

La primera aparición del protocolo de multiplexación FDMA en la telefonía móvil, fue en los equipos de telecomunicación de primera generación, en los años 1980, momento en el que la transmisión de datos era de muy baja calidad, y no existía ningún tipo de seguridad en esta.

2.3.2. Segunda generación

No fue hasta la década de los 90, cuando las redes de telefonía móvil pasaron de analógicas a digitales. Este cambio, y con él la aparición de la telefonía digital, vino de la mano de la red de segunda generación (2G), que se basaba principalmente en el estándar GSM (Global System for Mobile Communication).

La implementación de esta generación, aparte de pasar de un sistema analógico a uno digital, trajo consigo un aumento en la calidad y seguridad de las llamadas de voz, además de añadir una serie de funcionalidades. De las cuales las más importantes son las siguientes:

- Procesado digital de la voz.
- SMS.
- Roaming Internacional.
- Retención, transferencia y bloqueo de llamadas.
- Transmisión de datos (aunque con una muy baja velocidad).
- Autenticación y cifrado de datos.

La asignación del ancho de banda para las llamadas en las redes de segunda generación, se basaba en el protocolo de multiplexación TDMA, y la tecnología de acceso CDMA, y aumentó la velocidad de transmisión de datos con respecto a su predecesora, hasta una velocidad de 9,6 kbps.

Con respecto a la topología de red, que es lo que más importancia tiene para este trabajo, se utilizaba la correspondiente al estándar GSM. En esta segunda generación, el fronthaul de las redes estaba formado por varias estaciones de transmisión base (BTS, del inglés 'base transceiver station'), a las que los usuarios se conectaban a través de sus dispositivos móviles, que a su vez se encontraban conectadas a un controlador de estaciones base (BSC 'base station controller'), que era el encargado de ejecutar las distintas funciones necesarias para el funcionamiento de la red y los distintos protocolos implicados en el funcionamiento de las llamadas, así como de las distintas funcionalidades correspondientes a esta generación.

TDMA

El acceso múltiple por división de tiempo, o TDMA (del inglés Time Division Multiple Access), es una técnica de multiplexación utilizada para la transmisión de señales digitales, principalmente en las redes de segunda generación.

El funcionamiento de este protocolo TDMA, consistía en ocupar un único canal de transmisión (normalmente de una alta capacidad), por distintas fuentes de datos al mismo tiempo, de forma que se lograra un mayor aprovechamiento del ancho de banda disponible para la transmisión de datos. TDMA es una de las técnicas de multiplexación más utilizadas en la comunicación de datos.

La forma en la que se pueden conectar varias fuentes de datos utilizando un único canal mediante TDMA, consiste en dividir las unidades de información en ranuras (o slots) alternas de tiempo. Dicho de otra forma, las distintas fuentes de datos se van turnando para enviar su información a través del canal.

La ventaja que tiene este método de multiplexación es que permite enviar una cantidad mayor de datos, pues en muchos casos la capacidad del canal y su velocidad de transmisión era mucho mayor que la requerida por los dispositivos emisores y receptores y, de esta forma, al sumar los datos de varias fuentes en un mismo canal, su eficiencia y la cantidad de datos totales que se envían por el mismo aumenta mucho. También hay que decir que es el empleo de tecnología digital lo que permite el uso de esta técnica de multiplexación, pues a diferencia de la tecnología analógica que transformaba las ondas sonoras en electromagnéticas, al hacer la conversión a datos digitales, el receptor puede reinterpretarlos en el momento en que le lleguen, aunque no lo hagan de forma constante.

En el caso de la telefonía móvil, esta técnica de multiplexación se emplea dividiendo un único canal de frecuencia de radio en 6 u 8 slots, y asignando cada uno de estos slots a cada persona que realice una llamada, realizando el envío de datos de cada una de las llamadas en diferentes momentos, de forma que puedan emplear el mismo canal simultáneamente sin interferir entre sí. Esta técnica permite multiplicar el número de llamadas que se pueden realizar dentro de una celda de la red de telecomunicaciones sin necesidad de aumentar el ancho de banda al optimizar el aprovechamiento de los canales.

El estándar GSM, que es el empleado en las redes de segunda generación está basado en la técnica de multiplexación TDMA.

CDMA

La multiplexación por división de código, acceso múltiple por división de código o CDMA (del inglés Code Division Multiple Access) es un término genérico para varios métodos de multiplexación o control de acceso al medio basados en la tecnología de espectro expandido. Habitualmente se emplea en comunicaciones inalámbricas (por radiofrecuencia), aunque también puede usarse en sistemas de fibra óptica o de cable.

CDMA resuelve el problema del control de acceso al medio, empleando una tecnología de espectro expandido y un esquema especial de codificación, por el que a cada transmisor se le asigna un código único, escogido de forma que sea ortogonal respecto a los del resto de transmisores conectados al canal. De esta forma, el receptor capta las señales emitidas por todos los transmisores al mismo tiempo, pero gracias al esquema de codificación (que emplea códigos ortogonales entre sí) puede separar y seleccionar la señal de interés si conoce el código empleado a pesar que todas las señales compartan la misma frecuencia.

En CDMA, la señal se emite con un ancho de banda mucho mayor que el precisado por los datos a transmitir; por este motivo, la división por código es una técnica de acceso múltiple de espectro expandido. A los datos a transmitir simplemente se les aplica la función lógica XOR con el código de transmisión, que es único para ese usuario y se emite con un ancho de banda significativamente mayor que los datos.

Dentro del método de división de código, encontramos dos categorías distintas: CDMA síncrono (mediante códigos ortogonales) y asíncrono (mediante secuencias pseudoaleatorias).

El CDMA síncrono aprovecha las propiedades matemáticas de ortogonalidad entre vectores, cuyas coordenadas representan los datos a transmitir. La idea, es que el producto escalar de dos vectores ortogonales es 0, por lo que, aunque el receptor reciba una combinación lineal de varios

vectores (que sería la suma de las señales procedentes de varias fuentes) al mismo tiempo, si conoce el código de transmisión del usuario concreto cuyos datos quiere extraer, siempre podrá aislar sus datos de los del resto de usuarios, simplemente realizando el producto escalar de la señal recibida con el código de ese usuario. Pues al ser este ortogonal al de todos los demás, el resultado del producto de las demás señales será 0, aislando la señal de interés, y anulando el resto. Este método se podrá realizar con tantos usuarios como deseemos, siempre que existan códigos ortogonales suficientes para el número de usuarios deseado, lo que se logra incrementando la longitud del código.

De esta forma, siempre que cada usuario de CDMA síncrono emplea un código único para modular su señal, y los códigos de los usuarios en una misma zona sean ortogonales entre sí, como el producto escalar de estos es 0, tendrán una correlación cruzada igual a cero, y no provocarán interferencias entre sí.

Para que los sistemas CDMA síncronos puedan funcionar correctamente, es necesario que no haya un excesivo retardo en la llegada de las señales. En el caso de la red de radiotelefonía, esta condición no siempre se puede cumplir, debido a que los teléfonos móviles no se pueden coordinar con las estaciones de radio con demasiada precisión. El hecho de que podamos movernos con nuestro dispositivo hace que la señal enviada pueda encontrar obstáculos, y tardar más o menos tiempo en llegar hasta el receptor, lo que causa una variación en los retardos de llegada, lo que hace que este sistema CDMA síncrono pueda dejar de funcionar correctamente, por lo que es necesario buscar otra forma de enfocar el problema.

El efecto de esta movilidad de los terminales, y los distintos retardos de llegada que causa esta, provoca que sea matemáticamente imposible crear secuencias de codificación que sean ortogonales en cualquiera de los instantes aleatorios en los que puede llegar la señal al receptor. Para solucionar este problema, en los sistemas CDMA asíncronos, se emplean secuencias únicas "pseudoaleatorias" o de "pseudo-ruido", también conocidas como códigos PN. Estos códigos PN son secuencias binarias que parecen aleatorias, pero que pueden conocerse de forma determinista si el receptor así lo necesita, y se emplean para codificar y decodificar las señales de interés de los usuarios de CDMA asíncrono que están empleando un mismo canal (o frecuencia de transmisión), de la misma forma en la que se empleaban los códigos ortogonales en el sistema síncrono.

A diferencia de los códigos empleados en los sistemas CDMA síncronos, cuyo funcionamiento se basa en su ortogonalidad, las secuencias PN empleadas en CDMA asíncrono, funcionan debido a que no presentan correlación estadística, de forma que la suma de un gran número de estas secuencias resulta en lo que se denomina interferencia de acceso múltiple, o MAI (del inglés, multiple access interference), y pueden estimarse como un proceso gaussiano de ruido. Explicado de forma más sencilla, si las señales de todos los usuarios son recibidas con igual potencia, en el momento de utilizar el código para decodificar la señal de interés, las señales del resto de usuarios aparecerán como ruido en relación con esta señal de interés. Además, la MAI aumenta en proporción directa al número de usuarios (al generar cada uno un cierto ruido), lo que provoca que a mayor número de usuarios conectados simultáneamente, el receptor perciba un mayor nivel de ruido, y con ello una mayor interferencia en la señal recibida.

Por otra parte, el hecho de que las secuencias sean aparentemente aleatorias y de potencia distribuida en un ancho de banda relativamente amplio conlleva una ventaja adicional, que es el hecho de que estas señales son más difíciles de detectar en caso de que alguien intente captarlas, ya que se confunden con el ruido de fondo. Esta propiedad ha sido aprovechada durante el siglo XX en comunicaciones militares, puesto que aumentaba la seguridad de la transmisión.

Debido a que todos los usuarios conectados generan ruido sobre el resto de las señales, es muy importante controlar la potencia de emisión, puesto que a diferencia de los sistemas CDMA síncrono, TDMA o FDMA, que pueden (en principio), rechazar por completo las señales indeseadas (ya sea porque están codificadas con un código ortogonal que las anula, porque se envían en distinto

momento de tiempo, o se modulan en una frecuencia distinta), en los sistemas CDMA asíncrono es distinto. En estos sistemas solo se rechaza estas señales parcialmente, quedando un ruido de fondo en función del número de usuarios. Por este motivo, si alguna de las señales no deseadas se recibe con un nivel de potencia muy superior a la señal de interés, es posible que su nivel de ruido sea tan alto que imposibilite separarla de la señal de interés.

Para evitar este problema, los sistemas CDMA asíncronos incluyen un control de la potencia de todos los emisores, de forma que la potencia de todas las señales recibidas por el receptor sea aproximadamente la misma. En el caso de los sistemas de telefonía móvil, la estación de radio emplea un esquema de control de potencia por bucle cerrado, para controlar estrictamente la potencia con la que emite cada teléfono en cada momento.

Los sistemas de división por código o CDMA se emplean en múltiples sistemas de comunicación por radiofrecuencia, tanto de telefonía móvil, transmisión de datos o navegación por satélite.

2.3.3. Generación 2.5

Entre la segunda y la tercera generación existió una tecnología de red intermedia o 2.5G. Esta generación amplió alguno de los servicios de la segunda generación, entre los que destacaban:

- Acceso a aplicaciones en red para dispositivos a través del protocolo WAP.
- Mejora del servicio de mensajería instantánea SMS, utilizando EMS, que permitía la inclusión de melodías e iconos dentro del mensaje.
- Creación del servicio de mensajería multimedia (MMS), que permitía insertar imágenes, sonidos, vídeos y texto dentro del mensaje.
- Servicios Peer-to-peer (P2P) utilizando el protocolo IP, permitiendo la navegación por la red, o el acceso al correo electrónico entre otras cosas.

Esta generación se basó en la tecnología GPRS (General Packet Radio Service), que permitía velocidades de transmisión superiores a los 100 kbps, y al igual que su predecesora, las funciones en el fronthaul de la red seguían realizándose de forma centralizada.

WAP

El protocolo de aplicaciones inalámbricas, o WAP (del inglés, Wireless Application Protocol), es un estándar abierto internacional para aplicaciones que utilizan las comunicaciones inalámbricas, por ejemplo: acceso a servicios de Internet desde un teléfono móvil.

Se trata de la especificación de un entorno de aplicación y de un conjunto de protocolos de comunicaciones, cuyo objetivo es normalizar el modo en que los dispositivos inalámbricos se pueden utilizar para acceder a una serie de aplicaciones, como pueden ser el correo electrónico, o las noticias, entre otros.

El organismo que se encargaba de desarrollar el estándar WAP en sus inicios fue el WAP Forum, fundado por cuatro empresas del sector de las comunicaciones móviles: Sony Mobile Communications, Nokia, Motorola y Openwave. Desde 2002 el WAP Forum es parte de la Open Mobile Alliance (OMA), consorcio que se ocupa de la definición de diversas normas relacionadas con las comunicaciones móviles, entre las que se encuentran las normas WAP.

No fue sino el avance de la telefonía móvil, con la fabricación de móviles cada vez más livianos y potentes, lo que provocó que el número de dispositivos fuera cada vez mayor, lo que unido al aumento de sus capacidades, volvió muy interesante para los proveedores de servicios y contenidos disponer de un entorno normalizado que les permitiera ofrecer un mayor abanico de servicios a los usuarios de las redes móviles.

Y es en esta situación donde WAP cobra sentido, al definir un entorno de aplicación y una pila de protocolos para aplicaciones y servicios a través de terminales móviles. Puesto que WAP consiste en un conjunto de especificaciones, que permiten que los desarrolladores diseñen aplicaciones de interconexión para terminales móviles.

La tecnología WAP permite que los usuarios de estos dispositivos puedan acceder a servicios disponibles en Internet. Pero existían algunas consideraciones que era necesario seguir a la hora de diseñar estos servicios, basadas fundamentalmente a las características y capacidades de los terminales y la red de telefonía móvil, que provocaban que no fuera posible diseñar los servicios de la misma forma que se haría para un ordenador personal. Algunas de las cuales son:

- Los teléfonos móviles tienen una pantalla significativamente más pequeña que la de un ordenador personal.
- Sus teclados también son más limitados.
- Tienen más limitaciones de memoria, tanto en memoria RAM, como en capacidad de almacenamiento.
- Su procesador es menos potente.
- Por lo general, la red de telefonía móvil no permite la misma capacidad y velocidad de envío de datos que el internet por cable, aunque esta diferencia no es tan grande a partir de las redes de la tercera generación..

2.3.4. Tercera Generación

Después de las redes de generación 2.5, llegaron las redes de tercera generación o 3G. Estas dieron un salto importante tanto en el funcionamiento de las llamadas y el envío de datos, como en los servicios que ofrecían. Los servicios más destacables ofrecidos por esta generación eran principalmente: acceso a internet, servicios de banda ancha, roaming internacional, e interoperabilidad, además de los ya ofrecidos en las redes de las anteriores generaciones.

Además, la gran diferencia de esta generación era que permitía enviar datos a una velocidad mucho mayor, pasando de los 100 kbps de la generación 2.5, a un máximo de 2 Mbps, en ciertas redes y condiciones concretas. Este aumento de la velocidad de la conexión permitía la transferencia de datos y voz de forma simultánea desde un mismo dispositivo con una única conexión, así como la transmisión de vídeo a tiempo real, lo que permitió la aparición de servicios y aplicaciones de videollamada o televisión móvil, entre otros.

La diferencia de funcionamiento entre esta y las anteriores generaciones fue que en las redes de segunda generación, en el momento de realizar una llamada, al emplear el protocolo de multiplexación TDMA, se abría una línea de comunicación que era utilizada durante la duración de esta. De forma que había que asignar una capacidad de ancho de banda y un camino concreto por el que enviar los datos de la llamada.

En cambio, en la tercera generación, los datos a enviar se dividían en pequeños paquetes, que incluían la información relevante para poder ser reordenados e interpretados en el destino. Esto permitía repartir los canales de forma más eficiente, y enviar cada paquete por un camino distinto

si era necesario debido al estado de ocupación de la red, lo que hizo que la cantidad de información que podía enviarse ocupando un mismo ancho de banda aumentase considerablemente.

El estándar en el que están basadas las redes de tercera generación en Europa es UMTS (Universal Mobile Telecommunications System). La tecnología de acceso en la que se basa este estándar es WCDMA, y la red que emplea se llama UTRAN, y al igual que en sus predecesoras seguía centralizando las funciones de red en un Controlador de la Red Radio (RNC), que dirigía una serie de estaciones de radio. Los elementos que forman la red UTRAN son:

- El equipo de usuario (UE), que sería en este caso el teléfono móvil con el que los usuarios podían conectarse a la red siempre que se encontrasen dentro del rango de cobertura de esta. Este dispositivo debía estar preparado para soportar el estándar y los protocolos necesarios para la comunicación.
- La interfaz Uu, a través de la cual el UE se conecta a la red UTRAN.
- Los Nodos B, que son los responsables de la transmisión y recepción radio entre el terminal móvil y una o más celdas UMTS, creando y manteniendo el enlace con el UE, es decir, los que hacen la función de estaciones de radio.
- La interfaz Iu, que conecta los distintos elementos dentro de la red UTRAN.
- Y los RNC, a los que se encuentran conectados los distintos Nodos B de la red, que son los encargados de gestionar todos los recursos de radio (RRM), y de ejecutar las distintas funciones necesarias para realizar la transmisión de datos, así como de la gestión de movilidad, y de la encriptación de los datos antes de que sean enviados desde o hacia el terminal móvil.
- Por último, las RNCs de esta red de acceso, se encuentran conectadas mediante enlaces Iu a la Red Central (o Core Network), que es la red por la que viajan los paquetes para llegar de un usuario a otro

Antes de la llegada de la cuarta generación, se lanzaron dos estándares mejorados en la tercera generación, que fueron, primero High-Speed Packet Access (HSPA) una fusión de dos protocolos móviles, High Speed Downlink Packet Access (HSDPA) y High Speed Uplink Packet Access (HSUPA), que permitían velocidades de bajada y de subida de 14 y 7,2 Mbps respectivamente. Y después Evolved High Speed Packet Access (también conocido como HSPA+), una evolución de HSPA, que permitía velocidades teóricas de hasta 84 Mbps de bajada y 22 Mbps de subida.

HSPA

La tecnología HSPA, o High-Speed Packet Access, es la fusión de los protocolos HSDPA (High Speed Downlink Packet Access) y HSUPA (High Speed Uplink Packet Access), siendo su objetivo extender y mejorar el rendimiento de las redes de telefonía móvil de tercera generación, antes de la llegada de la cuarta, por lo que muchas veces es referida como 3.5G o 3.75G.

Según las especificaciones teóricas, las redes que empleaban HSPA, podían alcanzar velocidades de bajada de hasta 14,4 Mbits/s y hasta 5,76 Mbits/s de subida, dependiendo, como es de esperar, de la saturación de la red y su implementación. Además de esto, redujo la latencia con respecto a las redes de tercera generación, lo que permitía aumentar hasta 5 veces la capacidad del sistema para realizar descargas, y hasta 2 la de las subidas, en comparación con los protocolos WCDM que le precedían.

Para explicar el funcionamiento de esta tecnología con más detalle, la mejor forma de hacerlo es explicar por separado el funcionamiento y especificaciones de los dos protocolos que la conforman.

HSDPA

La tecnología HSDPA (High Speed Downlink Packet Access o Acceso descendente de paquetes a alta velocidad), también denominada 3.5G o 3G+, es la optimización de la tecnología espectral UMTS/WCDMA, y consiste en un canal compartido en el enlace descendente que mejora significativamente la capacidad máxima de transferencia de información, pudiéndose alcanzar tasas de bajada de hasta 14 Mbps. Esta tecnología permite unas capacidades de descarga promedio cercanas a 1 Mbps. Esta tecnología se emplea combinada con HSUPA, que en este caso es la encargada de gestionar el enlace ascendente.

HSDPA, es conocida como 3.5G, debido a que es la evolución de la tercera generación (3G) de tecnología móvil, siendo el paso previo antes de la cuarta generación (4G).

Esta tecnología es totalmente compatible con WCDMA, lo que permite que las aplicaciones desarrolladas para esa tecnología funcionen también con HSDPA. Además, la mayoría de proveedores UMTS dan soporte a HSDPA.

La idea de HSDPA es aprovechar al máximo las redes WCDMA, empleando toda la capacidad de estas para proveer servicios de banda ancha, con un aumento en la capacidad de datos móviles, y una tasa de transferencia de datos más elevada a la que alcanzaba su predecesora. De la misma forma que UMTS incrementa la eficiencia espectral (o los datos que es posible enviar con un mismo ancho de banda), con respecto a GPRS, HSDPA la aumenta en comparación de WCDMA. Este aumento en la cantidad de datos que podemos enviar, y el aumento de las velocidades que viene con ella, no solo permite emplear nuevos tipos de aplicaciones, sino también que el número de usuarios que pueden conectarse a una red simultáneamente también aumente. En este caso, HSDPA provee entre tres y cuatro veces más capacidad que WCDMA. En cuanto a la interfaz de las aplicaciones en tiempo real, tales como videoconferencia y juegos multijugador online, esta actualización de la tecnología reduce la latencia de la red (por debajo de los 100ms), lo que permite un mejor tiempo de respuesta, y con ello una mejor experiencia para los usuarios.

HSDPA puede alcanzar estas tan altas velocidades de transmisión, gracias al uso de modulación de mayor orden (Modulación de Amplitud en Cuadratura 16 - 16 QAM), codificación variable de errores y redundancia incremental, así como la introducción de nuevas y potentes técnicas tales como programación rápida. Además, HSDPA emplea un eficiente mecanismo de programación para determinar qué usuario obtendrá recursos, lo que permite que el reparto de la red sea más eficiente según las necesidades de cada uno. Aparte de todas estas ventajas, el empleo de un protocolo de enlace ascendente optimizado (HSUPA), que complementa a esta tecnología, y la instalación de receptores avanzados y antenas inteligentes, permite que las capacidades de HSDPA aumenten aún más.

Finalmente, comparte sus canales de más de alta velocidad entre los usuarios del mismo dominio de tiempo, lo que representa el enfoque más eficiente.

HSUPA

HSUPA (High-Speed Uplink Packet Access o Acceso ascendente de paquetes a alta velocidad) es un protocolo de acceso de datos para redes de telefonía móvil con alta tasa de transferencia de subida (de hasta 7,2 Mbit/s).

Generalmente es referido como generación 3.75 (3.75G) o 3.5G Plus, debido a que es una evolución de HSDPA (High-Speed Downlink Packet Access, Acceso descendente de paquetes a alta velocidad), nombrado popularmente como 3.5G. La tecnología HSUPA ofrece una mejora en la conexión de subida de las redes de tercera generación (UMTS/WCDMA).

HSUPA fue definido en Universal Mobile Telecommunications System Release 6 estándar publicado por 3GPP, como una tecnología que ofrecía una mejora sustancial en la velocidad para el tramo de subida, desde el terminal hacia la red.

HSDPA y HSUPA, ofrecen altas prestaciones de voz y datos, y permitirá la creación de un gran mercado de servicios IP multimedia móvil. HSUPA mejora las aplicaciones de datos avanzados persona a persona, con mayores y más simétricos ratios de datos, como el e-mail en el móvil y juegos en tiempo real con otro jugador. Y además las aplicaciones tradicionales de negocios, junto con muchas aplicaciones de consumidores, se beneficiarán del incremento de la velocidad de conexión.

Con HSUPA se prevé el uso de un canal dedicado en el enlace ascendente en el que se emplearán métodos similares a los empleados en HSDPA.

2.3.5. Cuarta Generación

Por último, antes de llegar a las redes 5G, que son las que se están instalando en la actualidad, tenemos la cuarta generación de tecnologías de telefonía móvil o 4G. Los objetivos principales de esta generación fueron proporcionar alta velocidad, calidad, capacidad, seguridad y servicios de bajo coste para servicios de voz y datos, multimedia e internet a través de IP.

Para que una red de telefonía pueda considerarse dentro de esta categoría, debe permitir una velocidad máxima de transmisión de datos de 100Mbps en una situación de alta movilidad y de 1 Gbps cuando se está parado o se tiene una movilidad baja.

EL estándar asociado con estas redes 4G es LTE (Long Term Evolution), el cuál incorporó velocidades de transmisión de datos mayores a las que nos encontrábamos en los estándares de la tercera generación, llegando hasta una velocidad superior a los 100 Mb/s, además de una significativa mejora en el rendimiento, que permitió mejorar la experiencia de uso de la red móvil. Al igual que sus predecesoras, su funcionamiento se basa en el envío de datos digitalizados a través de ondas de radio, y además tiene la ventaja de ser compatible con las redes GSM y HSPA, lo cuál evitaba los problemas de conexión al moverse entre las redes LTE que se iban incorporando y la red ya instalada, al permitir el traspaso de una a otra durante una llamada o transferencia de datos.

Además, la red LTE aumentaba los beneficios de las redes de tercera generación, al ofrecer varias ventajas con respecto a esta como son:

- Mayor capacidad, permitiendo realizar un mayor número de llamadas en la misma red.
- Mayores velocidades de datos, llevando la velocidad de conexión de los usuarios por encima de 100 Mbps
- Aplicaciones mejoradas de banda ancha móvil para teléfonos móviles, computadoras portátiles, videocámaras, cámaras y otros dispositivos electrónicos de consumo.

Aun con estas mejoras, y aunque el estándar LTE suele considerarse 4G, no entra realmente en esta categoría, puesto que no cumple los requisitos establecidos por la UIT, en características de velocidades pico de transmisión y eficiencia espectral.

Aun así, la arquitectura de las redes que consideramos de cuarta generación o 4G viene determinada por el estándar LTE. En este caso, la arquitectura de la red LTE presenta una serie de cambios que la diferencian bastante de sus predecesoras. Esta arquitectura evolucionada fue llamada Evolved UTRAN (E-UTRAN), y buscaba reducir el coste de los equipos y la complejidad de la red. Para esto, el nodo controlador de red (RNC), desaparece, y se crean un nuevo tipo de

estaciones de radio, denominadas nodos B evolucionados (eNB), que incorporan las funciones de control de recursos de radio, control de calidad de servicio y movilidad. De este modo, todos los eNB se encuentran conectados a través de una red IP, y se pasa de una arquitectura en la que las funciones de la red se encontraban centralizadas en los RNCs, a una en la que se encuentran distribuidas y se realizan en las propias unidades de radio eNB.

2.4. Redes 5G

Después de explicar la evolución de las redes de telecomunicaciones a lo largo de las distintas generaciones previas, llegamos a las redes de quinta generación o 5G que se están desplegando en gran parte del mundo en la actualidad, y van a seguir marcando el estándar en el que se basen las redes de telefonía móvil durante los próximos años. Este trabajo se desarrolla sobre las redes de esta quinta generación, que además de dar un gran salto con respecto a sus predecesoras en lo que a la capacidad y velocidad de la red respecta, presentan una asignación de las funciones de red completamente distinta, siendo este cambio la motivación fundamental del trabajo, y la funcionalidad de estas redes cuya problemática busca resolver y analizar la herramienta desarrollada.

La evolución más importante entre las de redes 5G y sus predecesoras es el aumento del ancho de banda soportado para la conexión, lo que deriva en una mayor velocidad de carga y descarga, hasta el punto de que se espera que estas redes 5G puedan llegar a alcanzar hasta 10gbps de velocidad, al menos en el papel. Debido a este incremento de la capacidad de las redes, se espera que puedan ser empleadas en un mayor número de dispositivos además de por los teléfonos móviles, ya sean ordenadores de escritorio o portátiles, o un gran número de dispositivos que en la actualidad funcionan en el internet de las cosas (IoT). Además, hay que tener en consideración que los antiguos teléfonos con conectividad 4G no serán compatibles con las redes 5G, por lo que los nuevos teléfonos que se han ido lanzando al mercado a lo largo del último año incluyen en la mayoría de los casos soporte específico para esta conexión 5G.

Además del aumento en la velocidad de conexión, las redes 5G traen consigo otra serie de mejoras o evoluciones, entre las que se encuentran:

- Menor latencia, permitiendo una conexión más fluida, con un delay que en algunos casos será inferior a los 5ms, lo que supone una conexión prácticamente a tiempo real.
- Mayor fiabilidad, evitando cortes en la conexión independientemente de la movilidad.
- Mayor rango de conexión, permitiendo que los usuarios se conecten desde una mayor variedad de lugares
- Mayor capacidad, permitiendo que haya un mayor número de dispositivos y usuarios conectados a la red simultáneamente.

Todas estas ventajas, se observarán en mayor medida conforme se vaya ampliando el despliegue de la red 5G, puesto que actualmente aún se encuentra en proceso, y permitirán que se avance en muchas aplicaciones y servicios, que van desde la conducción autónoma de vehículos, hasta la capacidad de descargar archivos muy grandes, como películas o juegos, en unos pocos segundos, pasando por un desarrollo inimaginable del internet de las cosas, que nos permitiría recopilar información a tiempo real del tráfico o el tiempo, entre muchas otras cosas, lo que nos permitirá dar solución con mayor rapidez a muchos de los problemas actuales.

Este aumento tan importante de la velocidad se basa en el empleo de ondas de alta frecuencia, pero al utilizarlas nos encontramos con el inconveniente de que tienen muy poco rango físico, por lo que la red 5G exige que se empleen muchas más celdas en comparación con las 4G, por lo que

necesitan una gran tarea de despliegue, motivo por el que el tiempo que se tardará en ofrecer la conectividad 5G en todas las ubicaciones y condiciones en las está programada, será mayor que lo necesario para implementar las anteriores generaciones de telefonía móvil.

Para minimizar este problema, las redes 5G operan en 3 bandas de frecuencias: baja, media y alta. Empleando cada una de estas las frecuencias de 600-850 MHz, 2.5-3.7GHz y 25-39 GHz respectivamente.

Por tanto, la red 5G estará constituida por hasta 3 tipos de celdas, cada una con un tipo de antena diferente. Estas antenas, proveerán diferentes relaciones entre velocidad de descarga vs distancias y área de servicio, siendo que las antenas de banda alta proveerán la mayor velocidad, pero tendrán el menor alcance, mientras que las de banda baja proveerán un gran alcance, manteniendo velocidades similares a las ofrecidas por la red 4G. Por otra parte, las antenas de banda media, que son las más utilizadas actualmente, ofrecen una mejora significativa en las velocidades con respecto a las antenas de la red 4G. Al mismo tiempo que mantienen un alcance de la conexión relativamente amplio, pudiendo cada torre de radiotelefonía proveer servicio a varios kilómetros de distancia.

De esta forma se puede proveer una conexión de calidad independientemente de las circunstancias y características del entorno, y es posible utilizar la red 5G aunque las antenas de banda alta aún necesiten un tiempo relativamente largo hasta terminar su despliegue. Además, hay que tener en cuenta que el inconveniente principal de esta banda es su problema para atravesar ciertos materiales, lo que unido a su corto alcance provocará que su uso se limite a lugares muy concurridos donde sea realmente necesaria esa gran velocidad de conexión.

El estándar en el que se basan las redes 5G fue publicado en el 3GPP, popularizando esta definición desde finales de 2018, mientras que las características y capacidades que debe poseer una red para ser considerada oficialmente de 5G, fueron especificadas por la ITU en el documento IMT-2020.

Con esto ya tenemos una definición relativamente completa de en qué consisten las redes 5G, siendo este tipo de redes en las que se fundamenta este trabajo, y sobre las que se va a ejecutar la herramienta de análisis desarrollada.

Como se mencionó previamente, la diferencia principal con las anteriores generaciones, por la que este trabajo cobra interés, es su arquitectura de red. En las redes de segunda y tercera generación las funciones de red se encontraban completamente centralizadas en un controlador de red, que dirigía varias estaciones de radio de una misma zona. Mientras que en las de a cuarta generación se ejecutaban directamente en las estaciones de radio. Finalmente, las redes 5G utilizan una arquitectura conocida como Cloud RAN, o simplemente C-RAN, que rompe totalmente con el funcionamiento de sus predecesoras, al permitir que estas funciones de red se puedan ejecutar parcial o totalmente en la nube si las condiciones del Fronthaul lo permiten.

En este tipo de arquitectura, para poder mandar una función de red a la nube, es necesario que se cumplan una serie de condiciones con respecto al delay máximo y la capacidad de los enlaces que nos encontramos entre la estación de radiotelefonía y esa nube. Puesto que es necesario enviar una cantidad concreta de datos a procesar, a una velocidad concreta, para que puedan ejecutarse esas funciones sin que surjan problemas en la conexión y en los servicios de la red. Por este motivo, este tipo de arquitectura, permite que las funciones de red puedan ser ejecutadas tanto en la nube, como en las propias estaciones de radio, así como repartirlas entre una y la otra; dependiendo del estado y las condiciones en las que se encuentre el Fronthaul de la red.

Por tanto, el objetivo de esta herramienta es determinar cuál es el mejor reparto de estas funciones de red entre la nube y las estaciones de radio de una determinada red, en función de una serie de criterios que serán explicados más adelante. Todo esto teniendo en cuenta la influencia que pueda haber entre las funciones asignadas a una estación y las capacidades disponibles en los

enlaces que pueda querer utilizar otra distinta.

Con respecto a la implementación de estas redes, las primeras redes con capacidades de 5G fueron desarrolladas en 2017, pero no fue hasta 2019 que se instalaron las primeras redes comerciales utilizando este estándar. En la actualidad, las empresas de telecomunicaciones continúan investigando nuevas tecnologías para posteriores versiones, y se prevé que la expansión de las redes basadas en estos estándares continúe.

Capítulo 3

Problema a resolver

Como se explica en el anterior capítulo, las funciones de red en el fronthaul de una red de telefonía 5G, pueden realizarse tanto de forma distribuida en las propias estaciones de radio (RUs, del inglés, Radio Units), como centralizarse en una unidad en la nube, así como repartirse entre ambas posibilidades, centralizando solo una parte de las funciones de una RU. La cantidad de funciones centralizadas, es decir, el nivel de centralización, vendrá dado por las condiciones de la red, y según lo que sea más conveniente en cada situación concreta. Con el objetivo de simplificar y hacer más eficiente la asignación del nivel óptimo de centralización de todas las RUs de una red, se agrupan varias funciones de red con exigencias similares, en lo que desde ahora llamaremos niveles de split. Estos niveles de split vendrán determinados por la capacidad y delay máximo en el camino entre la RU desde la que queremos centralizar y la CU a la que serán asignadas las funciones centralizadas necesarias para realizar un determinado grupo de funciones.

Por este motivo, el objetivo de este trabajo, es resolver el problema de la centralización óptima de estas funciones, y para resolver este problema, es necesario seleccionar la mejor combinación de caminos y niveles de split para cada una de las estaciones de radio en una red dada, según la capacidad y retardo de los enlaces de esta, y teniendo en cuenta la influencia del tráfico generado por cada una de las RUs según su nivel de split asignado, sobre la ocupación de la red, y por tanto sobre los niveles de split que es posible asignar al resto de RUs.

Por otro lado, los criterios que el programa puede utilizar para determinar cuál es la mejor solución (o nivel de centralización) en cada caso, son completamente variables dentro de la implementación según las necesidades. En este caso, y para todo el análisis realizado, el programa sigue cuatro criterios diferenciados, y selecciona la mejor solución (caminos y niveles de split para todas las RUs de la red analizada) según cada uno de estos en cada ejecución. Estos criterios son:

- Mínimo retardo total entre los caminos seleccionados para todas las RUs de la red.
- Mínimo retardo máximo para el camino seleccionado desde cualquiera de las RUs.
- Máxima centralización de las funciones de red en las CUs, o dicho de otra forma, máximo nivel de split centralizado sumando todas las RUs.
- Máxima centralización mínima de cualquiera de las RUs, es decir, máximo nivel de split mínimo.

3.1. Presentación escenario

Como se ha explicado, el objetivo de este trabajo, y de la herramienta desarrollada, es determinar el mejor nivel de split para las distintas RUs que conforman el fronthaul de una cierta parte de la red de telefonía móvil, por lo que las redes que vamos a analizar se corresponden precisamente con ese fronthaul de la red, o dicho de otra forma, lo que va desde las estaciones de radio hasta los nodos de la nube.

Para esto, los escenarios que se van a introducir en la herramienta representan una versión simplificada de ese fronthaul, consistente en una red formada por varias unidades de radio (RU), varios nodos intermedios (INT), varias unidades conectadas con la nube (CU), y una serie de enlaces que conectan estos elementos.

En todas las ejecuciones realizadas durante la realización de este trabajo, estos enlaces son bidireccionales, y su capacidad y retardo se asignan de forma aleatoria, en función de la tecnología utilizada en los mismos. Pero estos parámetros se pueden configurar fácilmente para utilizar la herramienta en casos diversos, permitiendo emplear enlaces unidireccionales, o varias tecnologías de enlace distintas en una misma red.

Además, es importante recordar, como se ha explicado en el apartado anterior, que en el caso de que dos o más estaciones de radio envíen sus datos empleando un mismo enlace, la ocupación de este que pueda provocar el tráfico necesario para centralizar las funciones de red de una de las estaciones, puede afectar al nivel de centralización que puede ser asignado a las demás. Por lo que hay que tener en cuenta que los resultados obtenidos, dependerán tanto de las características de los enlaces y los caminos, como de la influencia entre distintas estaciones de radio.

En las redes analizadas, todas las RUs se encuentran conectadas por algún camino a al menos una CU, y la capacidad de los enlaces permite en cualquier caso la utilización del nivel de split más bajo para todas ellas, de forma que siempre se pueda encontrar un camino válido desde cualquiera de nuestras RUs a alguna CU.

Antes de pasar a explicar el funcionamiento de la herramienta, solo queda presentar el escenario de ejemplo, que muestra de forma gráfica cómo están conformadas las redes que la herramienta va a analizar, y se utilizará para explicar cómo funcionan los distintos algoritmos que emplea la herramienta para determinar la mejor solución de centralización.

Este escenario es el mostrado en la Figura 3.1, y tiene 3 RUs, 2 CUs y 5 INTs, y cada RU tiene varios caminos posibles por los que llegar a una CU, por lo que el algoritmo tendrá que seleccionar el camino y el nivel de Split más adecuado para cada RU, para cada uno de los cuatro criterios seleccionados.

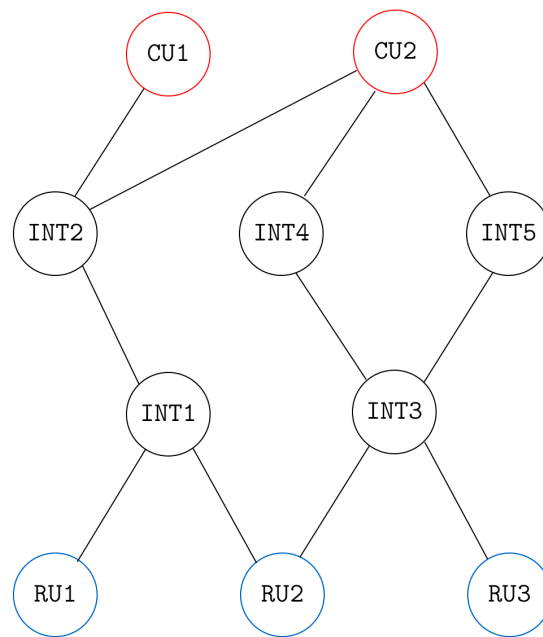


Figura 3.1: Red Ejemplo

Capítulo 4

Implementación

Una vez entendido el propósito de este desarrollo, qué es lo que se va a analizar, y cómo están formadas las redes sobre las que va a trabajar la herramienta, el siguiente paso es explicar el funcionamiento de la misma y las partes que la conforman, así como la forma en la que estas partes interaccionan entre sí, y cómo se introducen y modifican los datos de las redes a analizar.

Como ya se ha explicado, la idea base de este desarrollo es ejecutar una serie de algoritmos que permiten encontrar la mejor combinación de caminos y niveles de Split para cada una de las estaciones de radio de una red determinada, pero por encima de estos algoritmos, se encuentra una herramienta que permite controlar tanto la introducción y creación de redes a analizar, como la propia ejecución de estos algoritmos sobre las redes. Por este motivo, antes de explicar el funcionamiento de los algoritmos, es importante entender cómo funciona la herramienta de análisis, de la cuál he creado un diagrama de flujo que se encuentra en la Figura 4.1.

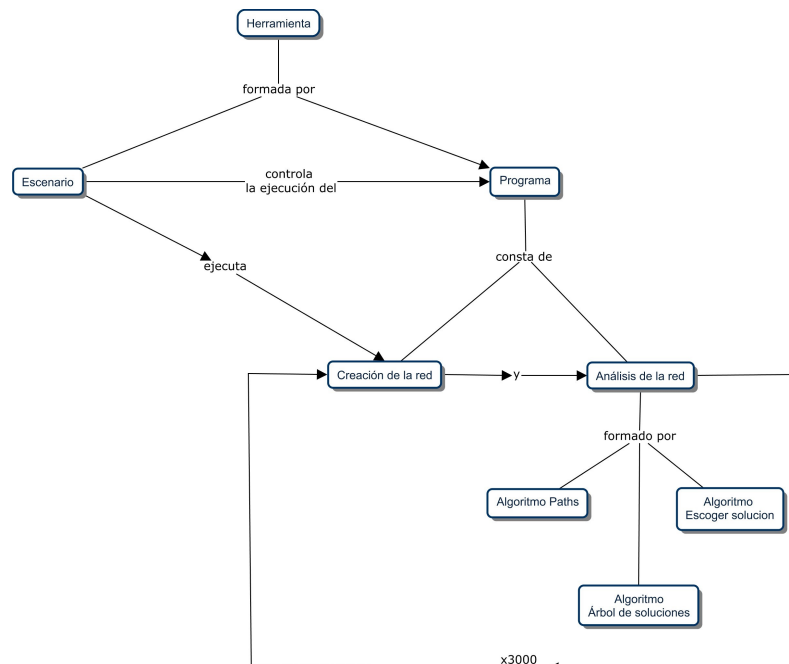


Figura 4.1: Herramienta Análisis

4.1. Herramienta

El funcionamiento general del programa consiste en analizar todos los caminos posibles por los que se puede llegar desde cada una de las estaciones de radio de una red a cada una de las estaciones de la nube, analizar todas las combinaciones posibles de caminos y niveles de centralización, y escoger la mejor combinación según unos determinados criterios. Pero hacer esto una sola vez en una única red no nos daría mucha información, puesto que la idea de este trabajo consiste en analizar cómo se comportan distintas redes según sus características, y como la mejor solución varía según las condiciones en las que se realice el análisis y los criterios empleados para escogerla, más que en encontrar una solución específica para un caso específico.

Por este motivo, la herramienta permite ejecutar el programa un número relativamente alto de veces consecutivas. En este caso, cada vez que se realiza una ejecución de la herramienta de análisis para una única topología de red, el programa se ejecuta un total de 3000 veces, encontrando la solución para las distintas combinaciones de tecnología de enlace, asignación de los nodos en la topología, y asignación aleatoria de las capacidades de los enlaces en función de la tecnología de enlace. Por tanto, lo que hace la herramienta, es encontrar la mejor solución para cada una de esas 3000 combinaciones en cada topología de red introducida.

Una vez entendida la idea de la herramienta, el siguiente paso es explicar las dos partes de las que se conforma, que son el escenario, y el propio programa.

A grandes rasgos, el escenario es el encargado de controlar la ejecución de las distintas partes del programa en el orden correcto, y gestionar los datos con los que trabaja la herramienta; mientras que el programa son las distintas partes que se ejecutan a la hora de encontrar las soluciones para una configuración de red concreta de las 3000 que se analizan para una topología.

Este programa consta de dos partes principales, que son la creación de la red a analizar, y la ejecución de los distintos algoritmos que permiten encontrar la solución correcta para esa red en concreto. En los siguientes apartados se explicarán estas partes más concretamente, pero lo primero es explicar en qué consiste el escenario.

4.2. Escenario

Como se ha explicado en la sección anterior, la herramienta que hemos desarrollado busca poder ser ejecutada un número relativamente alto de veces sobre una misma topología de red, cambiando ciertos parámetros y características. Concretamente, lo que varía entre las redes creadas para una misma topología es la tecnología de enlace empleada, el tipo del que es cada uno de los nodos, y las características de los enlaces según su tecnología.

Sabiendo esto, a partir de este momento, a cada una de las redes individuales creadas a partir de una topología de red determinada las llamaremos fotografías.

En el análisis realizado, la herramienta se ha ejecutado sobre tres topologías de red distintas, cada una de las ellas con un número de nodos distinto, y sobre cada una de estas topologías se han generado un total de 3000 fotografías, que se corresponden con todas las combinaciones resultado de multiplicar las 3 tecnologías de enlace empleadas, por 10 asignaciones de los nodos distintos por cada topología, y 100 variaciones de los enlaces por cada una de estas.

El escenario es, por tanto, la parte de la herramienta encargada de controlar la ejecución del programa sobre estas 3000 fotografías, indicando en cada caso al programa qué tecnología de enlace y qué asignación de los nodos está empleando en cada ejecución, y ejecutando las dos partes en el orden correcto.

Además de esto, el escenario se encarga de crear las instancias, tanto de las distintas acciones y algoritmos que conforman el programa, como de los distintos nodos que conforman la red, según el número de estos de cada tipo indicados en un fichero de configuración.

Por tanto, una vez entendido que es el escenario el que se encarga de ejecutar las dos partes del programa en el orden correcto para cada una de las fotografías a analizar, el siguiente paso es explicar cuáles son estas partes y en qué consisten.

4.3. Creación de la red

La primera parte es la creación de la red. Esta red se crea en función de la topología sobre la que estemos ejecutando la herramienta, y en función de las características correspondientes al número de la fotografía o de la iteración del programa en la que nos encontremos, pues tal y como se explicó antes, para cada topología de red se crean 3000 fotografías distintas, que deben crearse según los parámetros que toquen en cada caso.

Por tanto, la creación de la red se realiza teniendo en cuenta el caso en el que nos encontremos, y consta de tres partes diferenciadas, que son las siguientes:

La primera parte consiste en asignar la organización de los nodos en la topología, es decir, determinar qué nodos de la topología van a ser de cada tipo. El número de nodos de cada tipo será el indicado en el fichero de configuración del escenario, y a partir de esos datos, se determinará mediante el uso de una variable aleatoria qué nodos de la topología serán RUs y cuáles CUs, quedando el resto como nodos intermedios.

Para asignar esta organización de los nodos, se emplea la misma semilla para crear la variable aleatoria durante 100 ejecuciones, lo que permite que la asignación de los nodos sea la misma para estas 100 ejecuciones entre las que el programa cambia únicamente los valores de los enlaces, tal y como se indicó en la sección anterior.

El segundo paso consiste en asignar, empleando también una variable aleatoria, la capacidad y la longitud (que determinará el delay) de cada uno de los enlaces que conforman la red, siempre teniendo en cuenta los parámetros válidos para la tecnología de enlace empleada. En el caso de que el análisis hubiera sido realizado empleando enlaces de distintas tecnologías en una misma fotografía, en este apartado sería en el que se determinase la tecnología de enlace asignada a cada enlace, pero al haber empleado una única tecnología en cada caso, y tener cada una de estas una capacidad fija, lo único que varía es la longitud, y con ella el delay de los enlaces de la red.

En este caso, a diferencia de en la asignación de los nodos, la semilla con la que se genera la variable aleatoria con la que se asignan las características de los enlaces, varía en cada una de las 100 ejecuciones para una misma asignación de los nodos, al ser esto lo que hace que esas 100 ejecuciones sean distintas, lo que permita obtener una muestra significativa de los posibles resultados.

Y, por último, el tercer y último paso, consiste en guardar cada uno de los enlaces junto con sus características (capacidad, delay, y nodo con el que conectan), en las instancias de los nodos de los que parten. Para esto, hay que tener en cuenta que todos los enlaces de la topología son bidireccionales, teniendo la misma capacidad y delay en ambos sentidos, por lo que, en el momento de crear un enlace en la instancia de un nodo con dirección a otro, deberá crearse un enlace con la misma capacidad y delay, y sentido inverso en el nodo al que iba dirigido el primero. El objetivo de este paso es que, a la hora de ejecutar los distintos algoritmos, estos puedan saber qué enlaces parten de cada uno de los nodos de la red, con qué otros nodos se conectan, y cuáles son sus características. Con esto la red ya estaría lista para pasar a la segunda parte del programa, desde la que se ejecutan los algoritmos.

4.4. Análisis de la red

Una vez creada la red asignando los distintos nodos y guardando los enlaces en las instancias de sus nodos correspondientes, ya tenemos la fotografía que queremos analizar, y con esto podemos pasar a explicar la segunda parte del programa. Esta parte se encarga precisamente de analizar la red, buscando la mejor solución de centralización para sus estaciones de radio, según los criterios seleccionados.

Para ello se emplean tres algoritmos cuyo funcionamiento se explica con todo detalle en el siguiente capítulo, y que cumplen las siguientes funciones:

- El primer algoritmo se encarga de encontrar y crear un listado de todos los caminos posibles que parten desde cada una de las estaciones de radio de la red a analizar y llegan hasta alguno de los nodos de la nube, guardando los enlaces y los nodos por los que pasan cada uno de estos caminos.
- El siguiente algoritmo se encarga de crear un árbol de soluciones a partir del listado de caminos. Es decir, lo que hace este algoritmo, es generar todas las posibles combinaciones de caminos y niveles de centralización para cada una de las estaciones de radio de la red a partir del listado de caminos obtenido en el anterior algoritmo.
- Por último, el tercer algoritmo es el encargado de encontrar la solución indicada en cada caso, es decir, selecciona la mejor combinación de camino y nivel de centralización para cada una de las estaciones de radio de la red a analizar, según los criterios indicados, y devuelve una serie de métricas de las soluciones escogidas.

4.5. Final de la herramienta

Una vez ejecutado el programa para las 3000 posibles redes dentro de una topología, se guardan los datos de interés, tanto de estas redes, como de las soluciones seleccionadas por los algoritmos, en unos archivos de texto, en los cuales también están indicados la tecnología de enlace utilizada, y la configuración de los nodos a la que se corresponden cada uno de los datos. De forma que posteriormente se puedan analizar varias métricas de estas soluciones.

Capítulo 5

Explicación de los algoritmos

Tras explicar en el anterior capítulo el funcionamiento general de la herramienta de análisis y en qué consiste el programa que esta ejecuta, en este capítulo se explicará de forma más detallada la funcionalidad y la ejecución de cada uno de los algoritmos que conforman la acción encargada de analizar y extraer los resultados de cada una de las redes generadas sobre la topología introducida en la herramienta. Esta acción funciona ejecutando sobre la fotografía generada (o la red generada por la acción anterior del programa), tres algoritmos de forma ordenada, el funcionamiento de cada uno de los cuales se explica en este capítulo con todo lujo de detalles.

5.1. Algoritmo Paths

El primer algoritmo es el encargado de buscar todos los caminos posibles desde cada una de las estaciones de radio a alguno de los nodos de la nube. El funcionamiento de este algoritmo se basa en un algoritmo de búsqueda en profundidad o DFS (del inglés Depth First Search), con un par de restricciones para evitar caminos inútiles y/o redundantes.

A grandes rasgos, lo que hace este paso es ejecutar el algoritmo de búsqueda tantas veces como estaciones de radio haya en nuestra red. Cada vez que se ejecuta parte de una estación de radio, mira la lista de enlaces que parten desde ella (que como se explicó en el anterior capítulo, se guarda en la instancia de ese nodo durante la creación de la red), escoge uno válido, y salta al nodo al que vaya dirigido. En el momento de hacer el salto, guarda ese nodo en la lista de nodos que conforman el camino actual, y la capacidad y el delay del enlace en una variable temporal que se encarga de guardar los datos de todos los enlaces del camino que estamos recorriendo. Entonces, mira en la lista de enlaces del nodo al que ha saltado, busca otro enlace por el que saltar desde este, y vuelve a repetir el proceso.

Para evitar que se formen bucles, antes de escoger un enlace por el que saltar, el algoritmo siempre comprueba la lista de nodos que conforman el camino actual antes de hacer un salto, de forma que nunca salte a un nodo por el que ya pasa el camino y para no alargarlo innecesariamente recorriendo enlaces de más, ni formar bucles. De esta forma el algoritmo va saltando de un nodo a otro cumpliendo esta condición, hasta que se de uno de los siguientes casos:

- El primer caso, es llegar a un nodo desde el que no quedan saltos posibles, momento en el que el algoritmo retrocede al nodo anterior del camino, y borra el nodo desde el que ha retrocedido de lista de nodos que forman el camino y su capacidad y delay de la variable de datos del camino. Una vez en el nodo anterior del camino, comprueba si este tiene otros

posibles enlaces por los que saltar, y sigue con el algoritmo.

- La segunda posibilidad es llegar a uno de los nodos de la nube. Momento en el que el algoritmo guarda el camino recorrido para llegar hasta él junto con todos sus datos en una lista de caminos, y de igual forma que en el anterior caso, una vez hecho esto, retrocede al nodo anterior del camino. Puesto que no tiene ningún sentido buscar un camino más largo desde ese punto si ya ha llegado hasta la nube.

El funcionamiento de este algoritmo se basa en una función recursiva, es decir, que se llama a sí misma. Al llamar a esta función se le pasa como parámetro el nodo del que partimos, y lo que hace es leer su lista de enlaces, de forma que en el momento en el que encuentra uno por el que puede saltar, se llama a sí misma pasándose como parámetro el nodo al que quiere saltar, de forma que esta nueva instancia de la función pasa a leer la lista de enlaces del nuevo nodo en el que nos encontramos.

De este modo, cuando la función ha leído todos los enlaces de un nodo y no ha podido encontrar ninguno por el que saltar, o cuando completa un camino al llegar a un nodo de la nube y no desea seguir avanzando, no tiene más que salir de esa instancia de la función, retrocediendo a la instancia que se corresponde con nodo anterior desde el que saltó hasta ese, y borrando automáticamente los datos del camino que ha visto que no es válido, pues se guardan en una variable que se encuentra en la propia instancia de la función.

En este caso, el algoritmo ejecuta esta función recursiva una vez por cada una de las estaciones de radio de la red a analizar, y cada una de estas ejecuciones se termina en el momento en el que la función retrocede hasta llegar de vuelta a la estación de radio inicial, y no le quedan más caminos por leer que salgan de esta, de forma que se sale de esa primera instancia de la función, momento en el que ya se han recorrido y guardado todos los caminos válidos que llegan a los nodos de la nube desde la estación de radio desde la que se hizo la llamada.

5.1.1. Funcionamiento del algoritmo

En esta sección explicaré el funcionamiento del algoritmo sobre una pequeña red de ejemplo, haciendo uso de un esquema ilustrativo de los distintos pasos y acciones este ejecuta para obtener los resultados deseados. En este caso, la red de ejemplo tiene 3 estaciones de radio o RUs (del inglés, Radio Units), 2 nodos en la nube o CUs (del inglés, Cloud Units), y 5 nodos intermedios o INT, que se encuentran conectados entre sí por un total de 11 enlaces. Pero aun siendo una red muy sencilla, la cantidad de saltos que tiene que dar el algoritmo para analizarla completamente es bastante grande, por lo que únicamente explicaré su funcionamiento para la primera estación de radio, y se mostrarán las imágenes más relevantes para seguir la explicación.

Como ya se ha explicado antes, el algoritmo funciona mediante una función recursiva, que se va llamando a sí misma para ir avanzando por los distintos nodos de la red, por lo que el hecho de que en el esquema está marcado uno de los nodos de la red, significa que la instancia de la función en la que nos encontramos en ese momento ha recibido como parámetro de entrada ese nodo para que busque si existe algún salto válido en la lista de enlaces que parten de él. Una vez entendido esto, creo se puede empezar con la explicación del funcionamiento.

El algoritmo comienza en una de las estaciones de radio, en este caso la RU1, como se puede ver en la figura 5.1a. Para ello, llama a la función iterativa dándole como parámetro de entrada esa estación de radio. Entonces, busca entre los enlaces que están en la lista de ese nodo (en este caso, los enlaces que salen de la RU1), y decide por cual saltar, al tiempo que guarda el nodo en el que nos encontrábamos en una lista temporal que guarda el camino actual, tanto para poder guardarlo cuando lleguemos a uno de los nodos de la nube, como para que el algoritmo pueda

comprobar esa lista antes de saltar a un nodo que ya es parte del camino, evitando que se formen bucles.

En este caso, solo hay un enlace válido, que nos lleva a el nodo INT1, por lo que, como se puede ver en la imagen 5.1b, lo que hace es llamar de nuevo a la función pasándole como parámetros tanto el nombre del nodo al que está saltando, como el camino recorrido hasta ese momento. Entonces, la función busca de nuevo en la lista de enlaces del nodo actual, escogiendo saltar al INT2, y volviendo a guardar y a enviar los datos del camino.

En el nodo INT2 se repite la misma operación, saltando hasta la CU1, momento en el que hemos llegado a la primera unidad de la nube, y por tanto, encontrado el primer camino válido para la RU1. Esto lo podemos ver en la imagen 5.1c. Debido a que el camino ya es válido, y sería absurdo buscar un camino más largo que siguiera desde ahí, en el caso de encontrarnos en una CU, como en ese momento, lo que hace el algoritmo es guardar el camino como válido, y salirse de esa instancia de la función, de forma que, al encontrarnos en una función iterativa, al volver a la instancia anterior de la función, retrocede al nodo anterior del camino, que era el INT2, como se puede ver en la imagen 5.1d.

Esa función ya había empezado a buscar en la lista de enlaces de ese camino, con lo que ya había saltado por el primer enlace de la lista del nodo INT2, por lo que buscará el segundo enlace en la misma, que en este caso va a la CU2, un salto válido, por lo que salta por él. En este momento nos encontramos en la imagen 5.2a, que de nuevo es un nodo en la nube, por lo que de igual forma que en el anterior caso, guarda el camino como válido y vuelve a retroceder, situándose de nuevo en el INT2.

La función sigue buscando en la lista de enlaces de ese nodo, quedando solo uno, que es el que salta al nodo INT1, nodo que ya es parte del camino, por lo que el algoritmo no salta a él. En ese momento, al no quedar más posibles enlaces desde el nodo INT2, el algoritmo sale de esa instancia de la función, retrocediendo de igual forma al nodo INT1. En principio esto puede resultar confuso, puesto que acabo de decir que el algoritmo no ha querido saltar al INT1, pero la diferencia radica en que en un caso se estaba alargando el camino de forma redundante al avanzar a un nodo por el que ya hemos pasado, y en este caso no se está alargando el camino sino retrocediendo por él. Es decir, en caso de haber saltado, el camino sería RU1-INT1-INT2-INT1, mientras que al retroceder, el camino actual es simplemente RU1-INT1, y el nodo intermedio 2 se está borrando del camino.

En este punto nos encontramos en la figura 5.2b. Y para entender esta explicación que acabo de dar, en el esquema, los saltos marcados en rojo, son los únicos que forman parte de los caminos que serán guardados, mientras que los saltos en azul, solo indican la salida de una de las instancias de la función iterativa, eliminando un nodo y el salto dado para llegar hasta él, del camino actual.

Desde el nodo intermedio 1, el algoritmo salta a RU2, INT3, INT4 y finalmente a CU2, para encontrar el siguiente camino válido, como se puede ver en la imagen 5.2c. Y desde este nodo, tras guardar el camino, retrocede hasta el nodo INT3, pasando por INT4, tal y como podemos observar en la figura 5.2d, puesto que no hay más enlaces válidos que partan de INT4.

Desde ahí salta a INT5 y a CU2 de nuevo, encontrando y guardando el cuarto y último camino válido para esta estación de radio 1, mostrado en la imagen 5.2e.

Aunque en este momento sea fácil ver desde fuera que ya no quedan más caminos válidos, el algoritmo sigue funcionando, puesto que su forma de llegar a esta conclusión no es otra que agotar todos los posibles saltos e ir saliendo de las distintas llamadas de la función iterativa hasta volver a la primera instancia de esta que partía de la estación de radio desde la que estamos buscando los caminos, y quedarse sin enlaces por los que saltar desde ella.

Por este motivo, lo que hace el algoritmo en este momento es volver a retroceder hasta el nodo intermedio 3, desde el que salta por el enlace válido que le queda, que le lleva a la RU3 tal y como

se ve en la imagen 5.2f. Nodo desde el que no le quedan más enlaces válidos por los que saltar, pues su única conexión es con el nodo INT3, que es desde donde ha saltado, por lo que retrocede al INT3, desde ahí a RU2, INT1, y finalmente RU1, tal y como se puede ver en la figura 5.3a. Una vez en este nodo, al no quedarle ningún enlace por el que saltar desde esa estación de radio, sale de esa instancia inicial de la función iterativa, y termina la ejecución del algoritmo para para la primera estación de radio. En ese momento, el algoritmo guarda todos los caminos encontrados en la instancia de esa estación de radio, y vuelve a llamar a la función iterativa partiendo de la siguiente estación de radio, que es la RU2, para buscar los caminos válidos desde esta, pero creo que con la explicación del funcionamiento para la primera estación de radio se ha debido entender el algoritmo.

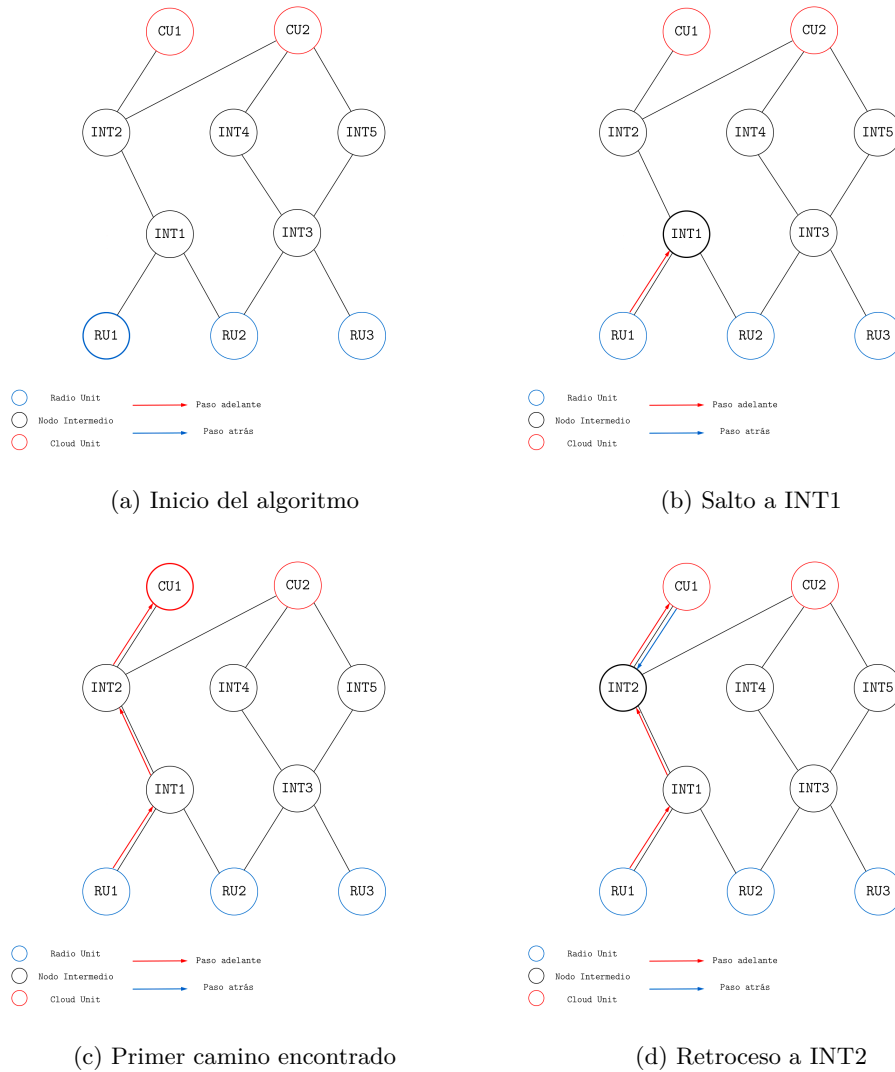
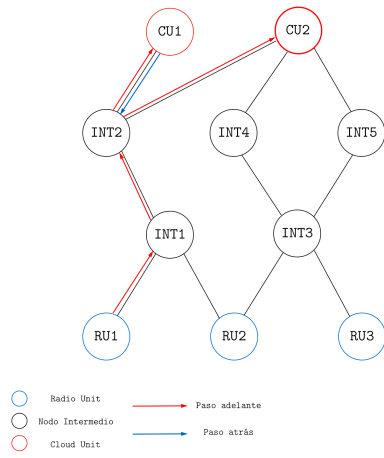
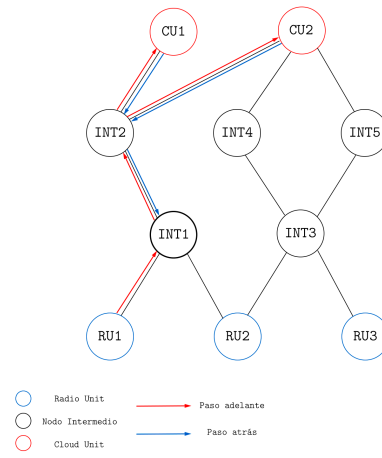


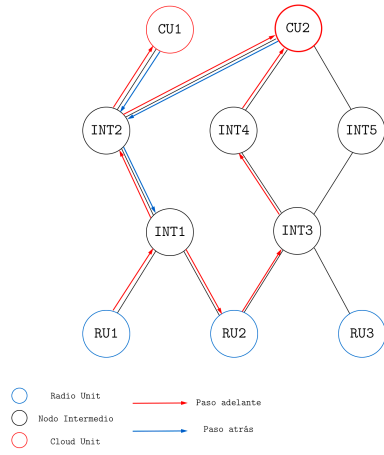
Figura 5.1: Algoritmo Paths



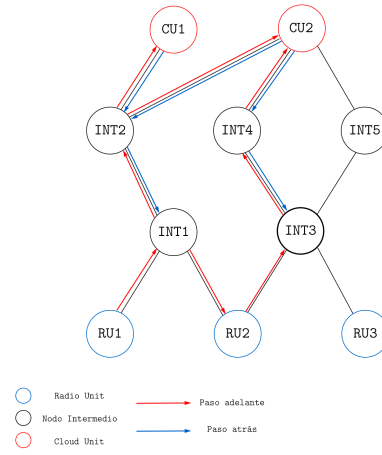
(a) Segundo camino encontrado



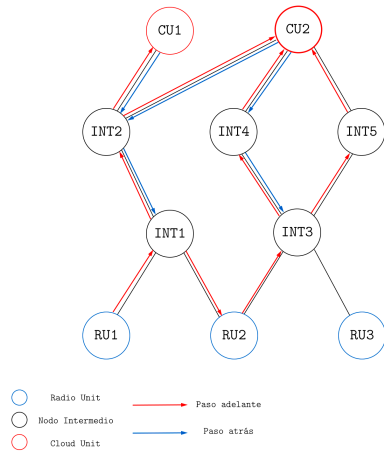
(b) Retroceso hasta INT1



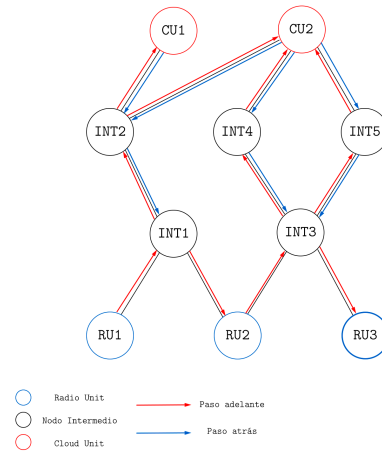
(c) Tercer camino encontrado



(d) Retroceso a INT3

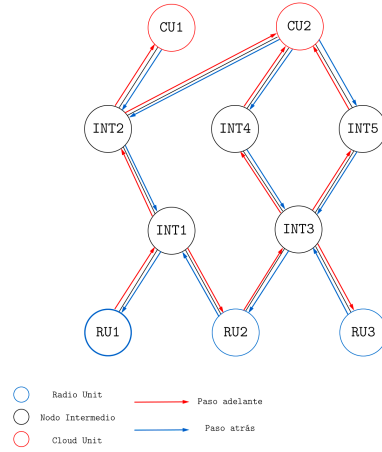


(e) Cuarto camino encontrado



(f) Sin caminos en RU3

Figura 5.2: Algoritmo Paths



(a) Final del algoritmo

Figura 5.3: Algoritmo Paths

5.2. Algoritmo Arbol Soluciones

Una vez encontrados todos los caminos posibles, el objetivo del siguiente algoritmo es crear un árbol con todas las posibles combinaciones de caminos y niveles de centralización para todas las estaciones de radio de la red.

Para entender el funcionamiento de este algoritmo, es importante entender el concepto del nivel de split. Como se ha explicado en el anterior capítulo, este algoritmo se encarga de crear todas las combinaciones de niveles de centralización y caminos para todas las estaciones de radio de la red.

El nivel de centralización es la cantidad de funciones de red que se envían a la nube, para que sean ejecutadas allí, y para que esto pueda hacerse, el camino desde la estación de radio a la nube debe cumplir unas ciertas condiciones en forma de capacidad o velocidad y delay del mismo, debido a que estas funciones necesitan enviar una cierta cantidad de datos con un retardo máximo.

Debido a que el número de funciones distintas que se podrían centralizar es relativamente alto, no tiene sentido comprobarlas una a una, pues las combinaciones serían prácticamente infinitas a pocas estaciones de radio y caminos que haya en la red. Por este motivo, las funciones a centralizar se organizan en niveles de split, que son un conjunto de funciones con necesidades similares, de forma que a la hora de determinar si la red tiene la capacidad necesaria para poder centralizarlas, se comprueban todas las funciones que pertenecen al mismo nivel de Split simultáneamente. De esta forma, el número de combinaciones desciende, al limitarse a comprobar un número más pequeño de niveles de split, en lugar de todas las funciones una a una.

Por tanto, en adelante llamaremos nivel de Split al conjunto de funciones que se están centralizando en una estación de radio por un camino concreto.

Una vez explicado este concepto, el funcionamiento del algoritmo consiste en buscar las combinaciones de caminos y niveles de split asignables a las distintas estaciones de radio. Para hacerlo, lo que hace es escoger la primera estación de radio, y su primer camino guardado, y comprobar qué niveles de split se se pueden centralizar por ese camino en función de las capacidades de este, que guardamos durante la creación del mismo en el algoritmo paths, teniendo en cuenta todos los enlaces que lo conforman. Una vez tiene los niveles de Split asignables a este camino, escoge otro camino para la segunda estación de radio, y comprueba a su vez qué niveles de split se le

pueden asignar, teniendo en cuenta la ocupación de los enlaces provocada por el nivel de split que asignado a la primera estación de radio.

Para poder hacer esto, el algoritmo va guardando el estado de la red conforme va creando y avanzando por el árbol, es decir, que si por ejemplo para asignar un determinado nivel de Split a la estación de radio número 1 por un camino, se ha empleado toda la capacidad de un enlace de la red, antes de comprobar que niveles de split se pueden asignar a la estación de radio número 2, se habrá restado la capacidad empleada por el Split asignado a la estación de radio 1 a la capacidad disponible en ese enlace, de forma que al comprobar el Split asignable para la estación de radio 2 la capacidad disponible se tendrá en cuenta que ese enlace tiene una capacidad disponible en ese momento de 0. A grandes rasgos, esto significa que la forma de tener en cuenta el efecto de la centralización de una estación de radio en las demás, consiste en restar las necesidades de capacidad de su Split asignado a la capacidad de los enlaces que forman el camino seleccionado para esta.

De igual forma que en el anterior algoritmo, este también funciona mediante una función recursiva que se va llamando a sí misma para avanzar por el árbol, y vuelve a la anterior instancia de la función para retroceder, lo que permite guardar este estado de la red en una variable de cada llamada de la función, que se envía actualizada al llamarse a si misma.

En este caso, el algoritmo retrocede cuando llega a una combinación que no permite asignar siquiera el nivel de centralización más pequeño a la estación de radio que toque en ese momento, o cuando ha asignado un camino y un nivel de split válidos a todas las estaciones de radio de la red, momento en el que guarda tanto esta como las combinaciones con los niveles de split inferiores al asignado para esa última estación y camino (pues estos también serán válidos), y retrocede.

El algoritmo acaba cuando ha creado el árbol completo, es decir, cuando ha guardado en el mismo todas las combinaciones posibles de caminos y niveles de Split para todas las estaciones de radio de la red. Cada una de estas combinaciones será una de las ramas del árbol de soluciones entre las cuales el algoritmo deberá seleccionar la mejor en el último algoritmo.

El funcionamiento de este algoritmo es bastante complicado, por lo que será más fácil entenderlo al explicarlo gráficamente sobre una red de ejemplo.

5.2.1. Funcionamiento del algoritmo

Este algoritmo es mucho más complejo que el anterior, generando un árbol de una magnitud exponencialmente mayor al anterior, por lo que, en este caso, para explicar el funcionamiento del mismo, emplearé una red aún más sencilla, y que sirva para explicar las interacciones que pueden generar las distintas estaciones de radio y sus niveles de centralización entre sí.

Como se ha explicado, este algoritmo tiene como propósito generar un árbol de soluciones en el que aparezcan todas las posibles combinaciones posibles de caminos y niveles de centralización para todas las estaciones de radio de la red que estamos analizando, de forma que el siguiente algoritmo pueda seleccionar la mejor entre estas posibles soluciones según el criterio o criterios determinados.

En este caso, a diferencia del algoritmo Paths, que se limitaba a encontrar los caminos desde cada una de las estaciones de radio de la red por separado, este algoritmo busca los niveles de centralización que se pueden asignar a las distintas estaciones de radio simultáneamente, por lo que el nivel de centralización asignado a alguna ellas, podría afectar al nivel de centralización asignable a las demás, en el caso de que los caminos seleccionados por estas compartieran alguno de sus enlaces, y la capacidad de alguno de ellos fuera limitada.

Por otro lado, para explicar el funcionamiento de este algoritmo, voy a simplificar el sistema de

niveles de split, reduciendo el número de estos con fines explicativos, y cambiando sus exigencias de capacidad y retardo a números más sencillos, de forma que sea más fácil ver claramente las interacciones que puedan suceder durante la asignación de los mismos por el algoritmo, y el funcionamiento de este.

Más concretamente, durante esta explicación, los niveles de centralización que el algoritmo podrá asignarle a una estación de radio los hemos reducido a dos, que llamaremos Split A y Split B. El Split A centraliza un menor número de funciones, y para funcionar exige un retardo total en el camino a la estación de radio de 15 o menos, y una capacidad de 1 o más. El Split B, en cambio, centraliza un mayor número de funciones, y por ello sus exigencias son mayores, necesitando para poder ser asignado, que el camino provoque un retardo en el envío de los datos de 10 o menos, y ocupando una capacidad de 2 en los enlaces que emplea.

Como ya he dicho antes, estos datos no tienen nada que ver con los niveles de split reales, y han sido asignados de forma arbitraria, con el único objetivo de lograr explicar el funcionamiento del programa de forma sencilla, y es por ese motivo, que ni siquiera indicaré unas unidades de medida para los retardos y capacidades necesarias. El algoritmo real emplea un número de niveles de split mayor, con las exigencias reales de las distintas funciones a centralizar, lo que aumenta bastante su complejidad con respecto al esquema que vamos a ir explicando en este caso.

La red empleada para la explicación, como se puede ver en la imagen 5.4 tiene únicamente dos estaciones de radio, dos nodos intermedios y un nodo en la nube, y todos sus enlaces tienen 2 de capacidad y provocan un retardo de 5 cada uno en el envío de los datos, de forma que por cualquier enlace de la red se pueda mandar el mayor nivel de split de los que acabamos de explicar

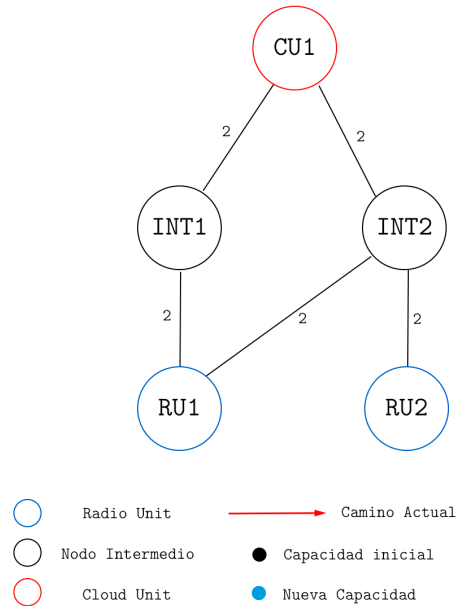


Figura 5.4: Red Explicacion

Hay que tener en cuenta que el tamaño del árbol de soluciones generado por este algoritmo crece exponencialmente con el número de estaciones de radio de la red, en función del número de caminos válidos que tengamos desde cada una de estas estaciones de radio y del número de niveles de split que podamos asignar. Para que se entienda la magnitud del crecimiento, en la red empleada para explicar el funcionamiento del anterior algoritmo, siendo una red realmente sencilla, con solo 3 estaciones de radio con 4 caminos válidos desde cada una, y considerando que se pudieran asignar 5 niveles de split distintos, podríamos tener un total de 8000 soluciones distintas en función de las características de los enlaces de la red. Obviamente sería imposible dibujar un esquema explicativo de ese árbol, y es ese el motivo por el que he reducido la red aún más, y los niveles de split asignables a solamente dos.

Para poder explicar el funcionamiento del algoritmo sobre esta nueva red de ejemplo, hay que considerar que se ha ejecutado sobre ella el algoritmo Paths, explicado en el anterior capítulo, obteniendo como resultado que desde cada una de las estaciones de radio de esta red salen dos caminos que llegan al nodo de la nube, tal y como se puede observar en la figura 5.5.

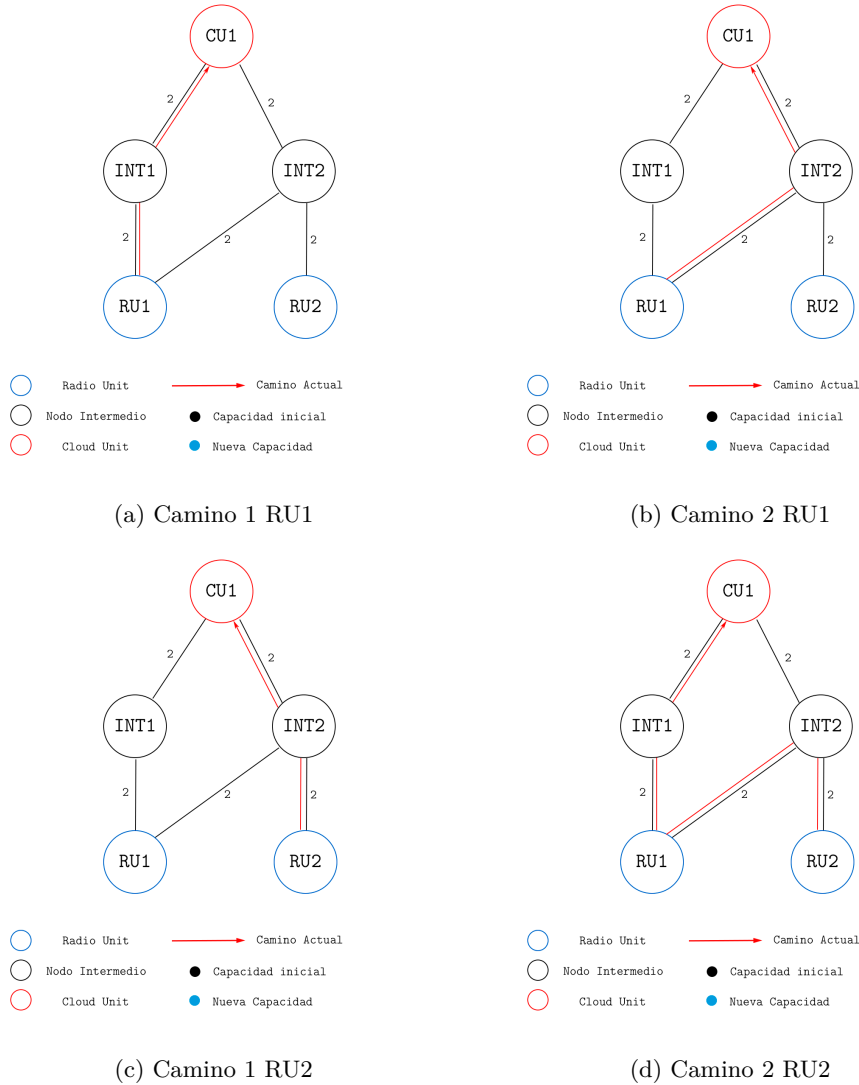


Figura 5.5: Caminos válidos

Una vez explicados los motivos por los que ha sido necesario simplificar la red de ejemplo y los niveles de split que se van a emplear durante la explicación, es posible comenzar a explicar el funcionamiento de este algoritmo. El algoritmo empieza a crear el árbol por la primera de las estaciones de radio, que en este caso es la RU1, seleccionando el primer camino de esta, que es el mostrado en la imagen 5.5a, y asignándole el nivel de Split A, que es el más pequeño de los dos posibles. Entonces, la función comprueba que ese camino cumpla las condiciones necesarias para poder asignarle ese nivel de split a la estación de radio utilizándolo.

A cada combinación de un camino y un nivel de split válidos para una estación de radio determinada, lo llamaremos una decisión para esa estación de radio, de forma que cada rama del árbol se irá formando tomando "decisiones" para las distintas estaciones de radio de la red, y conformará una solución válida si ha sido posible asignar un camino y un nivel de split válidos a todas ellas. El esquema que representa el árbol muestra precisamente estas combinaciones de caminos y niveles de split, marcando cuales son válidos y cuales no según las condiciones de la red. Por tanto, para la explicación de este nivel de split, mostraré tanto el esquema de la red, con los caminos empleados y las distintas actualizaciones de la capacidad de los enlaces, como el árbol que va creando el algoritmo conforme va comprobando las distintas decisiones para cada estación de radio, y llegando a soluciones válidas.

Como he dicho antes, el algoritmo empieza en la RU1, probando el nivel de split A en el camino 1, tal y como se puede ver en la imagen 5.6a. En este caso el Split A exige una capacidad mínima de los enlaces del camino de 1, y un retardo máximo a lo largo del mismo de 15, siendo que los enlaces que forman ese camino tienen ambos una capacidad de 2, y que al ser un camino formado por únicamente dos enlaces, el retardo generado es de 10, por lo que el camino y el nivel de split asignado serían válidos. Una vez el algoritmo los detecta como tal, genera un nuevo estado de la red, restando la ocupación generada por el nivel de split asignado a la capacidad de los enlaces por los que pasa el camino empleado, quedando un nuevo estado de la red tal y como se puede ver en la imagen 5.8a; y guarda el camino y nivel de split asignados a la RU1 en una variable temporal que indica la combinación de decisiones, o la solución actual.

Igual que el algoritmo Paths, este algoritmo también funciona mediante una función recursiva que se llama a sí misma, por lo que en este caso, una vez una asignación de camino se ha dado como válida, lo que hace el algoritmo es llamar a la función de nuevo, pasándole como parámetros la combinación de decisiones que tenemos hasta ese momento (que sería la rama en la que nos encontramos) y el estado de la red actualizado.

En este caso, el siguiente salto del algoritmo comprueba el primer camino para la RU2, que es el indicado en la figura 5.5c, y prueba el menor nivel de split para él, tal y como se puede ver en la figura 5.6b. Igual que en el caso anterior, es posible asignar el nivel de Split A a este camino, debido a que cumple las exigencias de retardo y capacidad necesarias para hacerlo, y de igual forma que en el anterior paso, genera el nuevo estado de la red, quedando la imagen que se puede ver en la imagen 5.8b, y guarda este nuevo camino y nivel de split en la variable de las decisiones actuales.

Si hubiera una tercera estación de radio en nuestra red, volvería a llamar a la función para buscar un camino y un nivel de split válido para esta, pero al no ser así, ya hemos encontrado un camino y nivel de split para todas las estaciones de radio de nuestra red, por lo que guardaría la combinación de soluciones que tenemos hasta ese momento en el árbol final, siendo esta la primera rama de ese árbol, tal como se puede ver en la imagen 5.8c. Entonces, retrocede en el árbol para poder seguir buscando desde el salto anterior, es decir, en este caso, borra de la variable temporal de decisiones el camino y el nivel de split asignados a la RU2, y partiendo del estado de la red que encontramos en la imagen 5.6a, comprueba si hay algún otro nivel de split que se pueda asignar al mismo camino de la RU2.

El otro nivel de split que se puede probar es el split B, que exige una capacidad de los enlaces

de 2, y un retardo no mayor de 10. Ese camino puede cumplirlas, de forma que esta combinación de decisiones que podemos ver en la imagen 5.6c también es válida, con lo que es guardada en el árbol de soluciones que en este momento es el mostrado en la imagen 5.8d. En este punto ya no quedan más posibles niveles de split para ese camino de la RU2, en esta rama del árbol puesto que ya han sido comprobados los dos que pueden ser asignados, por lo que tras retroceder de nuevo por el árbol, pasa a comprobar si hay algún nivel de split que se pueda asignar al camino 2 de la RU2, que es el que podemos ver en la figura 5.5d.

Como en el anterior camino, empieza comprobando el nivel de split más bajo, puesto que este es el que tiene menores exigencias de capacidad y retardo para funcionar. El camino 2 de la RU2 tiene un total de 4 enlaces, los cuales suman un total de 20 de retardo, y el nivel de Split A exige un retardo máximo de 15, por lo que no es posible asignárselo a este camino. Debido a que no es posible asignar el menor nivel de split, el algoritmo no comprueba el nivel de split mayor para ese camino, puesto que las exigencias de capacidad y retardo para que se pueda asignar este son aún mayores.

Durante el funcionamiento del algoritmo real con un mayor número de niveles de split, el algoritmo funciona de la misma forma. La función va comprobando a lo largo de cada rama del árbol si se pueden asignar los distintos niveles de split de menor a mayor, y en el momento en el que llega a uno que no puede asignar en una rama concreta del árbol, no sigue comprobando los niveles de split superiores al mismo, debido a que estos siempre van a tener las mismas o mayores exigencias, y sigue probando otros caminos para la estación de radio en la que se encuentre en caso de quedar alguno que no haya probado aún, o retrocede por la rama para probar otras combinaciones.

Una vez el algoritmo ha comprobado ambos caminos y sus posibles niveles de split para la RU2, retrocede a la RU1, y en este caso, comprueba si se puede asignar el nivel de split B al camino 1, que es en el que nos encontrábamos anteriormente. El camino cumple con las condiciones de retardo necesarias para asignarle ese nivel de split, tal y como se puede ver en la imagen 5.6d, por lo que lo da como válido, lo añade a la variable de la solución actual, que había borrado en el momento que retrocedió hasta la primera estación de radio, y vuelve a comprobar el nivel de split A para el camino 1 de la RU2. De igual forma que en las combinaciones con el nivel de split A en la RU1, al emplear un camino para la RU2 que no comparte enlaces con el camino empleado para la RU1, tanto el nivel de split A como el B vuelven a ser válidos, dando lugar a los estados de la red que se pueden observar en las imágenes 5.6e y 5.6f, por lo que guarda esas dos nuevas ramas de la red, teniendo un total de 4 soluciones válidas hasta este momento.

Para entender el funcionamiento de este algoritmo tengo que insistir en la idea de función recursiva. La función va comprobando cada una de las RUs de la red, llamándose a sí misma, y pasándose como parámetros tanto el estado de la red provocado por las estaciones anteriores a la actual, como la combinación de decisiones que llevamos hasta ese punto, que serían las que formarían la rama del árbol que posteriormente vamos a decidir si guardar como válida o no, en el caso de encontrar una solución posible para todas las estaciones de radio de la red. Y siguiendo la misma lógica, la forma de retroceder por la red no es más que salir de una llamada a la función, lo que nos sitúa en la anterior llamada, donde aún no se han pasado esos datos como parámetros, por lo que recuperaremos el anterior estado de la red, y aún no habremos añadido esa solución a la combinación anterior.

En este punto ya sabemos que el camino 2 de la RU2 no permite ningún nivel de split válido en ningún caso, puesto que el retardo que genera es mayor al exigido para asignar el menor nivel de split, por lo que es irrelevante el camino o el nivel de split asignados a la RU1. El punto en el que la interacción entre los caminos utilizados por distintas estaciones de radio cobra relevancia es la exigencia de capacidad de los niveles de split. Puesto que si hemos asignado un cierto nivel de centralización a una estación de radio, la ocupación de los enlaces que haga esta, afecta a la capacidad que tenemos disponible para centralizar las funciones de otras estaciones de radio. Por

este motivo, en adelante, no mencionaré en cada rama que el algoritmo determina que no se puede asignar ningún nivel de split al camino 2 de la RU2, pero hay que tener en cuenta que el algoritmo lo comprueba cada vez, pues no tiene ninguna forma de saber por qué lo rechazó en anteriores iteraciones de la función, y se limita a probar cada combinación de decisiones posible, y a rechazar las que no son válidas. Por lo que aunque no lo haya dibujado en el esquema, cada vez que el árbol comprueba las decisiones válidas para la estación de radio 2 en cada una de las ramas del árbol, también comprobará que el camino 2 no permite la asignación de ningún nivel de split debido a su retardo.

Explicado esto, ya hemos probado y guardado todas las ramas que parten del camino 1 de la primera estación de radio, teniendo un árbol con un total de 4 soluciones válidas hasta el momento, tal y como se puede ver en la figura 5.8e. El siguiente paso que toma el algoritmo es comprobar el siguiente camino de la RU1, en este caso el mostrado en la figura 5.5b, probando si puede asignar el nivel de split A a través el mismo. Este camino tiene una capacidad de 2, y un total de 2 enlaces que provocan un retardo de 10, por lo que es posible asignar ese nivel de split A. Por tanto, igual que en el resto de iteraciones, guarda el estado de la red tal como se puede ver en la imagen 5.7a, y lo manda como parámetro antes de probar el nivel de split A para el camino 1 de la RU2, encontrándonos con el árbol de la figura 5.8f. En este caso es donde ese estado de la red actualizado cobra relevancia, puesto que los caminos empleados por ambas estaciones de radio comparten un enlace, por la capacidad del mismo empleada por una de las centralizaciones se resta de la que puede emplear la otra. En este caso, si que se puede asignar el nivel de split A al camino 1 de la RU2, debido a que la capacidad restante en el enlace compartido es de 1, siendo justamente la exigida por ese nivel de split. De esta forma, tras asignar ese nivel de split, el estado de la red sería el mostrado en la imagen 5.7b.

Por tanto, tenemos otra rama válida, y el algoritmo pasa a comprobar el nivel de split B para ese mismo camino de la RU2, tal y como se ve en la figura 5.8g. Pero a diferencia del caso anterior, la capacidad exigida por este nivel de split es de 2, y el enlace compartido solo tiene una capacidad restante de 1, por lo que en el caso de asignarlo, el estado de la red resultante sería el mostrado en la imagen 5.7c, con un enlace con capacidad restante negativa, lo que obviamente no es posible, por lo que tal y como se puede ver en el esquema del árbol representado en la figura 5.8h, esta rama no es válida.

Por último, el algoritmo vuelve a retroceder, y comprueba que es posible asignar el nivel de split B para el camino 2 de la RU1. Entonces crea el nuevo estado de la red en función de la ocupación de esta asignación, tal y como se puede ver en la imagen 5.7d, y comprueba el nivel de split A para el camino 1 de la RU2, quedando un árbol como el mostrado en el esquema 5.9a. En este caso, la asignación del nivel de split B para la RU1 ha ocupado toda la capacidad del enlace compartido por los dos caminos, por lo que no es posible asignar siquiera el menor nivel de split a la segunda estación de radio por este camino. Con lo que el árbol resultante es el mostrado en la imagen 5.9b, y de igual forma que en el anterior caso, con esta asignación, el estado de la red resultante tendría un enlace con capacidad negativa, tal y como puede verse en la imagen 5.7e.

Como ya hemos explicado antes, si no se puede asignar el menor nivel de split para una estación de radio y camino determinado, el algoritmo no comprueba los niveles superiores, y puesto que el segundo camino para la RU2 tampoco permite la asignación de ningún nivel de split debido a su retardo, vemos que no hay ninguna combinación de soluciones válida que incluya la asignación del nivel de split B al camino 2 de la RU1, por lo que indicaremos que esa solución tampoco es válida aunque la red la permita individualmente. En este momento, ya no quedan más caminos ni niveles de split que probar para la RU1, por lo que el algoritmo habrá terminado su ejecución, quedando que el árbol final resultante es el mostrado en la figura 5.9c.

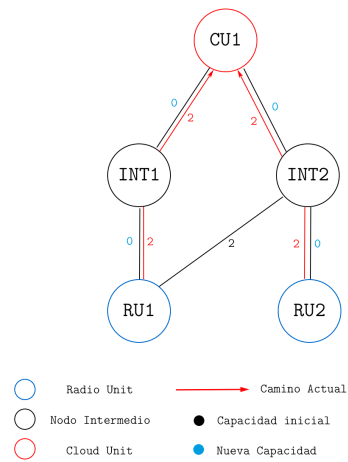
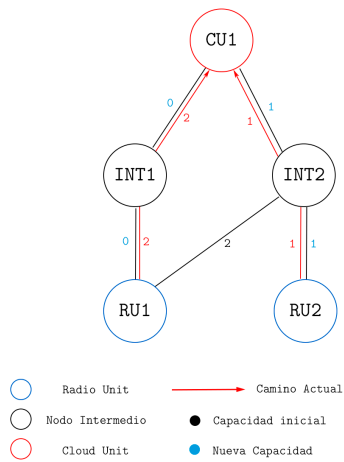
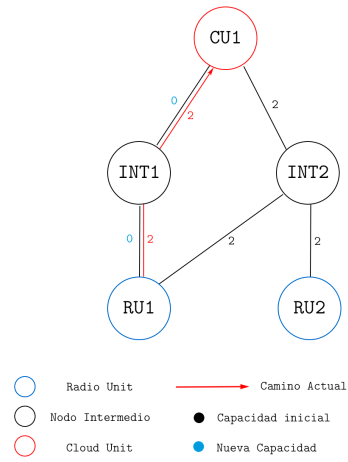
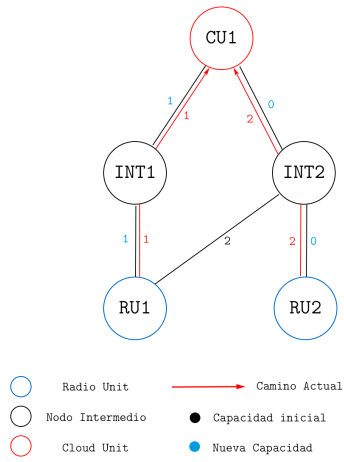
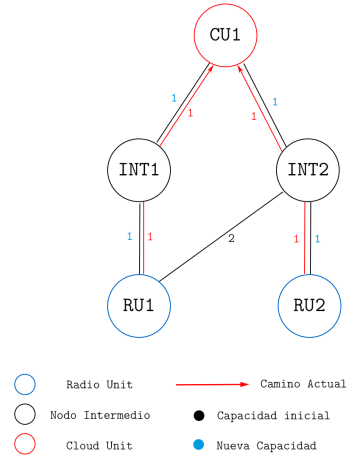
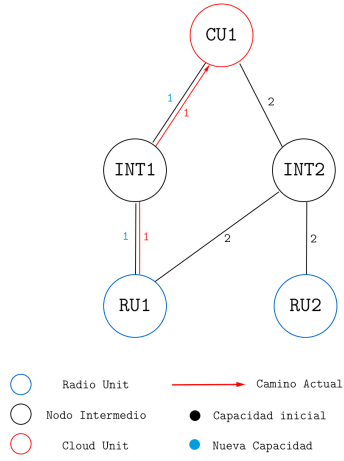


Figura 5.6: Red Algoritmo Árbol Soluciones

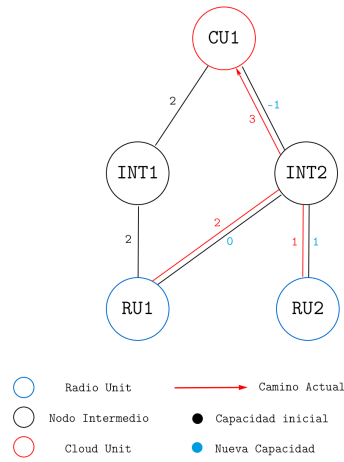
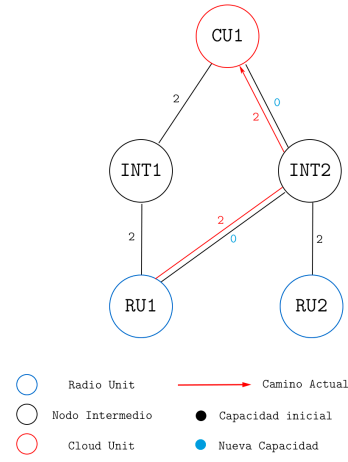
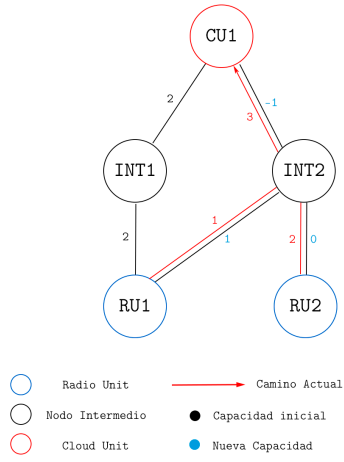
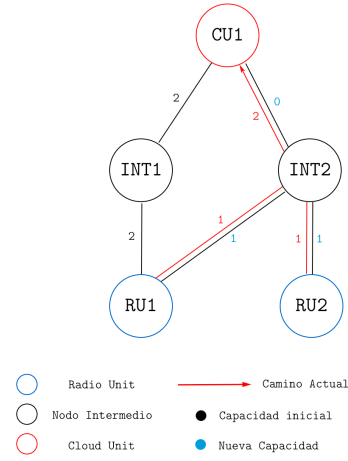
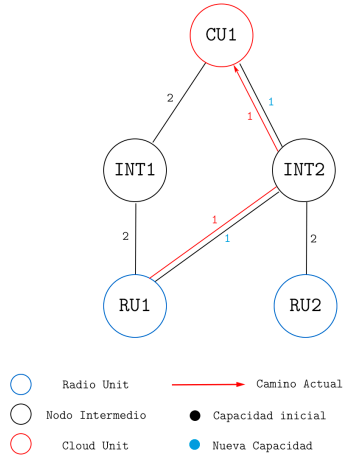
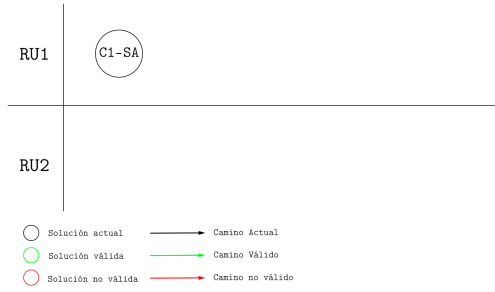
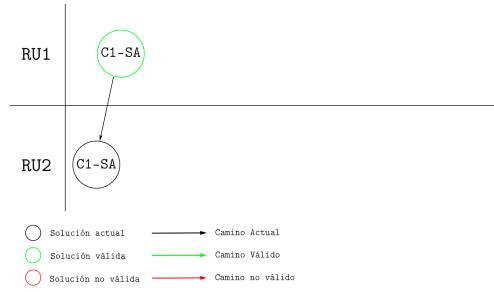


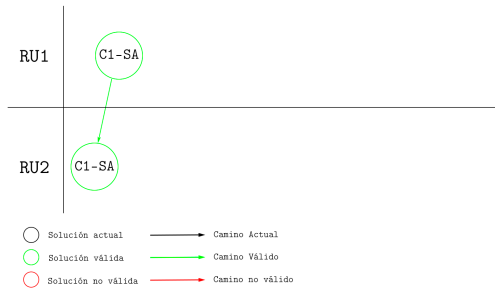
Figura 5.7: Red Algoritmo Árbol Soluciones



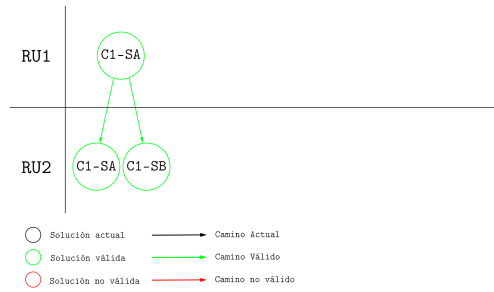
(a) Estado árbol 1



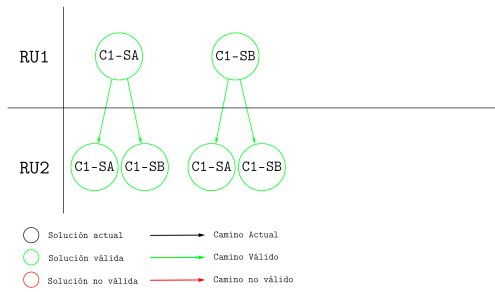
(b) Estado árbol 2



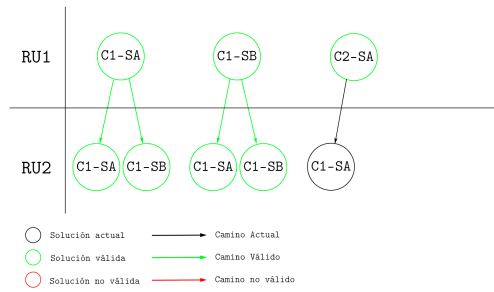
(c) Estado árbol 3



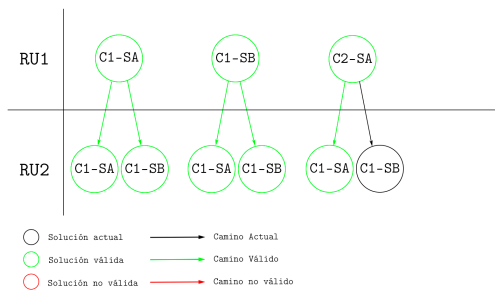
(d) Estado árbol 4



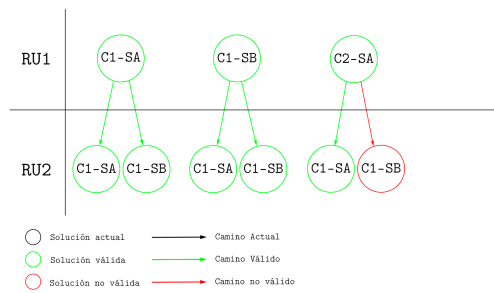
(e) Estado árbol 5



(f) Estado árbol 6



(g) Estado árbol 7



(h) Estado árbol 8

Figura 5.8: Árbol Soluciones

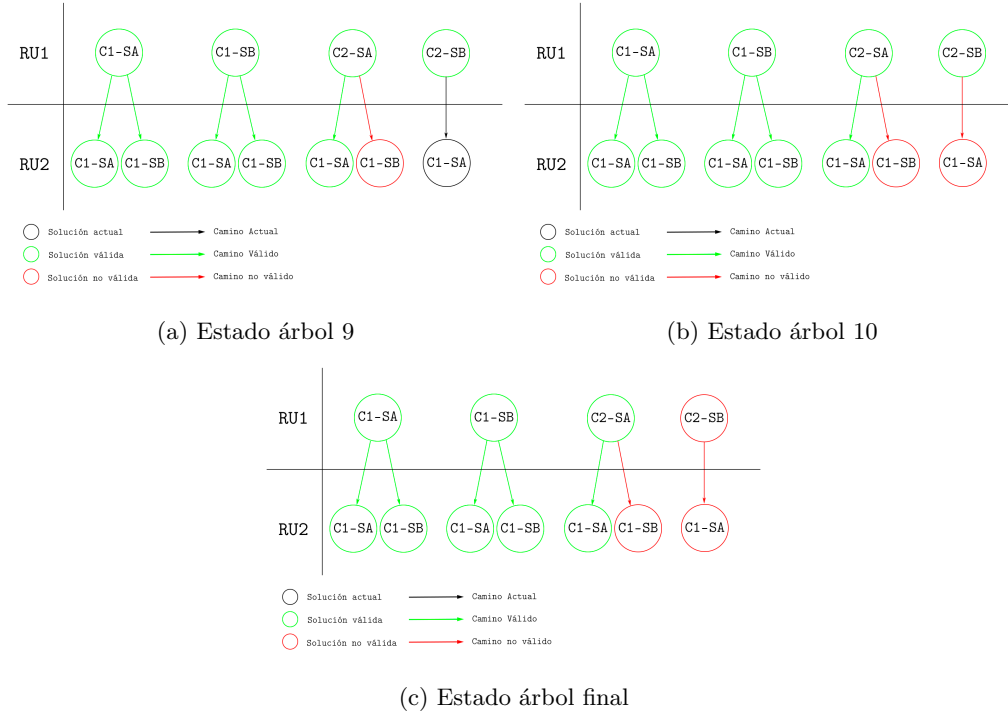


Figura 5.9: Árbol Soluciones

5.3. Algoritmo Escoger Solución

Una vez creado el árbol de soluciones, el último algoritmo se encarga de escoger la mejor de estas soluciones según varios criterios de selección. Su funcionamiento es bastante sencillo en comparación con los anteriores algoritmos, pues se limita a recorrer el árbol de soluciones, y comparar si la solución en la que está es mejor que las anteriores por las que ya ha pasado. El criterio que emplea el algoritmo en cada caso para escoger la mejor solución, es perfectamente configurable, pero en el caso de este trabajo, se ejecuta 4 veces, una con cada uno de los siguientes criterios de selección:

- Mínimo retardo total entre los caminos seleccionados para todas las RUs de la red.
- Mínimo retardo máximo para el camino seleccionado desde cualquiera de las RUs.
- Máxima centralización de las funciones de red en la nube o, dicho de otra forma, máximo nivel de split centralizado sumando todas las RUs de la red.
- Máxima centralización mínima de cualquiera de las RUs, es decir, máximo nivel de split mínimo.

Lo que hace en cada ejecución del algoritmo, es ir comparando todas las posibles soluciones una a una combinando los datos de todas las decisiones que conforman una solución. Para hacer esto, parte del inicio de la primera rama del árbol, creando unas variables en las que va a guardar los datos de la solución actual, a las que da un valor inicial de cero, lee el delay y el nivel de split asignados a la primera estación de radio, y se lo suma a las variables, al mismo tiempo que actualiza los datos de máximo delay y mínima centralización de la solución, y a continuación

sigue avanzando por el árbol, y sumando y guardando los datos correspondientes a las distintas decisiones que conforman la solución, hasta llegar al final de esta rama.

De igual forma que los anteriores algoritmos, este también recorre el árbol mediante una función recursiva, lo que permite que, al retroceder por una rama del árbol, se recuperen los datos que había antes de saltar a ese siguiente nodo, pues cada instancia de la función guarda los datos hasta llegar hasta ella en sus variables. Cuando llega al final de una rama, ya ha guardado los datos totales de la solución, y lo que hace es compararlos con los mejores datos según el criterio escogido en cada caso. La primera rama se guarda como mejor solución automáticamente, y en las siguientes, compara si según el criterio seleccionado, esa solución es mejor que la anterior mejor, caso en el cual la sustituye. En caso de que dos soluciones sean iguales según uno de los criterios, la herramienta utiliza un segundo o tercer criterio para desempatar, y si aun así siguen siendo iguales, simplemente mantiene la solución que había sido considerada como mejor antes.

El orden de desempate según el criterio seleccionado para cada caso es el siguiente:

- En caso de empate cuando el criterio seleccionado es el mínimo retardo total, desempatará con el mínimo retardo máximo, y por último con la centralización total.
- Si empata en el de mínimo retardo máximo, empleará el retardo total, y después la centralización total si siguen manteniéndose igualados.
- Con el de máxima centralización total, desempatará con la centralización mínima, y el retardo total.
- Y, por último, utilizando el criterio de máxima centralización mínima, el desempate se hará con la centralización total, y si no, con el retardo total.

En el caso de la máxima centralización, para calcular ese dato, puesto que no se puede considerar con exactitud si es mejor una estación de radio con nivel de split 4, y otra con nivel 1, o dos con split 2, lo que hace el programa es darle un valor a cada nivel de split, que ha sido decidido de antemano según las funciones que nos permite centralizar, y sumar la puntuación de todos los niveles de split de la solución, para poder determinar que combinación es mejor.

Para cada uno de los criterios, el algoritmo termina su ejecución en el momento en el que ha recorrido el árbol de soluciones completamente, momento en el que escribe en un fichero de soluciones los datos de la solución seleccionada.

5.3.1. Funcionamiento del algoritmo

Ahora, por si no ha quedado completamente claro el funcionamiento de este algoritmo, en esta sección se explicará de forma gráfica la ejecución del algoritmo sobre el árbol de soluciones generado por el algoritmo anterior, empleando uno de los 4 criterios de selección, lo que se correspondería con una única llamada al algoritmo. Mientras que, en el caso del programa real, la acción encargada de ir ejecutando los distintos algoritmos llama a este algoritmo un total de 4 veces, pasando como parámetro en cada una de esas llamadas el criterio que debe de ser empleado durante la ejecución del mismo para, de esta forma, poder extraer por separado la mejor solución según cada uno de los criterios.

Además, es importante tener en cuenta con respecto a esta explicación que, tal y como se explicó antes, en el caso de que dos o más soluciones o ramas del árbol fueran iguales según el criterio seleccionado, el programa emplearía el resto de criterios en un orden determinado para desempatar. Y en caso de que, tras probar todos los criterios, siguiera sin deshacerse el empate, simplemente se quedará con la rama que hubiera escogido primero como mejor solución.

En este caso, al emplear una red tan simple para la explicación, haber limitado los posibles niveles de split en la misma, y haber considerado que todos los enlaces provocan el mismo retardo en el envío de los datos, el resultado que obtendríamos en el caso de ejecutar el algoritmo sobre la misma, sería el mismo para los 4 posibles criterios, debido a que habría empates entre las ramas en los criterios relacionados con el retardo, puesto que los 4 caminos que finalmente son utilizados en las soluciones válidas para la red de ejemplo, causan todos el mismo retardo en el envío de datos, por lo que finalmente todos los criterios terminan desempataando en función de la centralización total.

Hay que tener en cuenta que esta situación también puede darse con redes reales, y sucede en algunos casos con las redes que ha analizado el programa, como se podrá ver en el capítulo que muestra los resultados obtenidos de la ejecución del mismo, aun siendo redes más complejas que esta red de ejemplo. Esto se debe tanto a que la variabilidad de los datos de los enlaces entre distintas ejecuciones del algoritmo no es excesivamente grande, a que las redes empleadas no tienen un gran número de enlaces, lo que provoca que el número de soluciones sea relativamente comedido, y en algunos casos, simplemente a que dos criterios distintos puedan llevar a la misma solución, lo que provoca que en muchos casos alguno de los criterios de lugar a empates entre ramas, y esto lleva a emplear un segundo criterio para desempatar. En cualquier caso, esto no es un problema, pues a nivel de análisis, que suceda no deja de ser información relevante con respecto al funcionamiento del programa y de las redes en si mismas, así como con respecto a los criterios de selección de la mejor solución.

Por este motivo, la explicación del funcionamiento de este algoritmo se hará empleando un único criterio de selección sobre la red de ejemplo, que en este caso será el de mínimo retardo máximo de los caminos correspondientes a las distintas decisiones que forman la rama del árbol, puesto que conozco de antemano que empleando este criterio se observarán varios empates entre ramas del árbol, de forma que tendré que emplear los dos criterios de desempate para llegar a la solución final, lo que me facilitará explicar la forma que tiene el programa de analizar y comparar los datos de cada uno de ellos.

El funcionamiento del programa, como se ha explicado antes, consiste en extraer los datos de las distintas ramas del árbol de soluciones, e ir seleccionando la mejor entre ellas. Para ello, simplemente se limita a guardar la mejor rama encontrada hasta el momento, y comparar las siguientes con esta. Para poder hacerlo, guarda en una variable todos parámetros a comparar de esta mejor solución, que en este caso son el delay máximo, el delay total y la centralización total. Pero antes de analizar la primera de las ramas, no hay una mejor solución con la que comparar esta, por lo que lo primero que hace el algoritmo antes de empezar el análisis, es inicializar estas tres variables con valores que incluso la peor de las combinaciones superaría, de forma que la primera solución se convierta en la mejor en todos los casos. Con los datos de la red de ejemplo, los valores con los que se inicializan estas variables son de 0 para la centralización total, y de 999 para el retardo máximo y el retardo total, ya que cualquier solución posible mejoraría esos valores, mientras que los valores de estas variables en el algoritmo real están basados en la máxima dimensión posible de las redes que vayamos a analizar.

Ahora que estas variables tienen valor, el algoritmo puede pasar a comparar la primera rama o combinación de decisiones con ellas. Para hacerlo, lo primero que tiene que hacer es extraer los datos de la misma, también en unas variables temporales, que de igual forma que en el anterior caso, también tienen que ser inicializadas, en este caso todas a 0. La primera rama del árbol es la que está indicada en la figura 5.10a, y para obtener los datos de la misma, tiene que ir sumando o guardando los datos de las distintas soluciones de la misma, que en este caso son el nivel de split A para el camino 1 de la RU1, y el nivel de split A para el camino 1 de la RU2.

Para hacer esto, el algoritmo va leyendo los datos de cada una de las decisiones de la rama, es decir, los caminos y niveles de split para cada una de las estaciones de la red y actualizando esas variables de la rama actual según se corresponda. La primera decisión que ha de analizar es la

correspondiente a la RU1, que en este caso, como he dicho antes y podemos observar en la imagen 5.10b es un nivel de split A, a través camino 1 de esta estación de radio, que es el que podemos observar en la imagen 5.5a. Entonces, el algoritmo lee el delay correspondiente a ese camino, (el cual, aunque no ha sido explicado, se calculaba durante la ejecución del algoritmo que generaba el árbol de soluciones, es decir, que en el momento de ir guardando las distintas decisiones de una rama, se calculaba adicionalmente el delay respectivo al camino asignado por cada una de estas y se guardaba en la instancia correspondiente del árbol), y lo suma al delay total de la rama, el cual, al ser esta la primera decisión de la rama, valía 0 hasta este momento.

Como podemos recordar del anterior capítulo, el camino 1 de la RU1, que es el mostrado en la imagen 5.5a, tiene dos enlaces, y tal como se explicó en ese capítulo, cada enlace de esta red de ejemplo simplificada provoca un retardo de 5, por lo que el retardo total de este camino será de 10, que sumado a 0 da un retardo total hasta el momento también de 10. En este punto, con la idea de ahorrar tiempo en el resto de la explicación, creo importante aclarar que todos los caminos que forman parte de las distintas ramas del árbol final, causan de igual forma un retardo de 10, ya que el único camino que no tenía dos enlaces era el camino 2 para la RU2, que se puede ver en la imagen 5.5d, el cuál vimos en el capítulo de la creación del árbol, que no permitía la asignación de ninguno de los niveles de split válidos en este caso.

Una vez calculado el retardo del camino, comprueba si el retardo de este es mayor al retardo máximo de los caminos de esa rama hasta el momento, comparando este valor de 10 con el de la variable temporal correspondiente, que en este momento era de 0, al ser esta la primera decisión y camino de la rama, por lo que actualiza el valor del mayor retardo de la solución, por el retardo del camino actual. Y, por último, llama a una función que extrae la ocupación asociada al nivel de split correspondiente a esta solución, que en este caso es el split A, cuya ocupación es 1, y se lo suma a la variable ocupación total, que ahora pasa a valer 1.

Con respecto a la suma de niveles de centralización, la idea es que el valor que le vamos a dar al nivel de split, para posteriormente sumarlo al total, irá determinado por la exigencia de capacidad del mismo. Es decir, que para hipotético nivel de centralización necesitase de 150 megas para funcionar, su valor a la hora de calcular cual es el mayor nivel de split, será de igual forma de 150. Esto es así, puesto que consideramos que, si el nivel de split exige una mayor capacidad de los enlaces para funcionar, es decir, necesita enviar una mayor cantidad de datos a través de la red, significa que la carga de trabajo necesaria para ejecutar esas funciones aumentará en la misma proporción.

De esta forma, ya hemos sumado los datos de la primera decisión de la primera rama, y el siguiente paso es hacerlo con la decisión correspondiente a la RU2, mostrada en la figura 5.10c, que igual que en el caso de la RU1, ha escogido su camino 1 y su split A, con lo que realizando el mismo proceso, y debido a que este camino también tiene dos enlaces, nos queda que el retardo total tras sumar ambas decisiones, es de 20, la ocupación total de 2, y el retardo máximo sigue siendo de 10, puesto que al comparar el de este camino con el valor anterior de esa variable, nos encontramos con que ambos son iguales, y por tanto el programa no lo actualiza.

Con esto, al no haber más estaciones de radio cuyos datos sumar a los que ya tenemos, ya habremos extraído todos los datos correspondientes a la primera solución, o rama del árbol, y lo siguiente que hará el algoritmo es compararlos con los de la mejor solución hasta el momento. Al ser esta la primera solución, obviamente no puede haber una mejor anterior, por lo que tendrá que ser seleccionada, y es ese motivo por el que las variables que guardan el mejor caso, habían sido inicializadas con datos que cualquier posible solución mejoraría.

Al estar ejecutando el algoritmo según el criterio de menor retardo máximo, lo que hace es comparar el valor de esa variable, que en este momento era de 999, con el correspondiente a la rama actual, que es de 10, comprobando si este segundo es menor. Al ser menor, esta es seleccionada como mejor solución hasta el momento, como se ve en la imagen 5.10d, y se actualizan tanto esta

como el resto de las variables de la mejor solución con los datos de esta solución, quedando que la mejor solución en este punto de la ejecución del programa tiene un nivel de centralización de 2, un delay total de 20, y un delay máximo de 10.

El siguiente paso, es comprobar la segunda rama. Tal y como se puede ver en la imagen 5.10e, esta solución tiene el mismo camino y nivel de split para la RU1, y el mismo camino, pero con el nivel de split B para la RU2. No creo necesario explicar paso a paso la extracción de los datos de la rama, pues el proceso es exactamente el mismo que con la primera, con la diferencia de que al cambiar un nivel de split, los resultados finales son que el delay total y el delay máximo son de 20 y de 10 respectivamente, igual que en el anterior caso, mientras que la centralización tiene un valor de 3.

De igual forma que en el caso anterior, lo que hace el algoritmo es comparar estos datos con los de la mejor solución hasta ese momento. Al estar ejecutando el algoritmo con el criterio de menor retardo máximo, primero compara esta variable, quedando que, tanto en la solución actual como en la mejor, vale 10. Al haber un empate, pasa a comparar el retardo máximo de todos los caminos, que de igual forma es de 20 para los dos casos, por lo que emplea el último criterio de desempate, que es la centralización total. En la solución actual, el valor es de 3, superando a la mejor hasta el momento, por lo que la solución actual pasa a ser la mejor solución, como vemos en la imagen 5.10f, y se actualizan los datos de las 3 variables con los de la solución actual.

Ahora pasamos a analizar la tercera rama del árbol, mostrada en la imagen 5.11a. Las decisiones de esta rama siguen empleando el camino 1 para ambas estaciones de radio, pero cambian en que en este caso se asigna el nivel de split B a la RU1, y el A a la RU2. De esta forma, los datos finales para ella son que el retardo total es de 20, el retardo máximo es de 10, y la centralización total de 3. Estos datos son exactamente los mismos que los de la mejor solución total, por lo que, tras comprobar los tres criterios, no logra llegar a un desempate, y como ya hemos indicado antes, el criterio que sigue el algoritmo, es que si una solución no llega a mejorar a la anterior mejor solución, se mantenga la que había sido indicada como tal con anterioridad. Por este motivo, pasa a analizar la cuarta rama, de la imagen 5.11b. La cual mantiene los caminos para ambas estaciones de radio, y el nivel de split de la RU1, pero aumenta el nivel de split de la RU2 al B. Eso hace que los valores de delay total y máximo sean los mismos que en el anterior caso, pero la centralización total para esta solución pase a ser de 2. Por lo que tras empatar en los dos primeros criterios, el criterio de máxima centralización causa que esta sea una mejor solución, con lo cual se actualizan los datos de las variables, y se guarda esta rama como la nueva mejor solución, como se indica en la imagen 5.11c.

Por último el algoritmo comprueba la rama de la figura 5.11d, que es la última combinación de decisiones que nos queda por comparar. En esta se emplea el camino 2 de la RU1, que de igual forma tiene un delay de 10, y el camino 1 de la RU2, mientras que asigna un nivel de split A a ambas estaciones de radio. En este caso, el retardo máximo y el total vuelven a empatar con la mejor solución, mientras que el nivel de centralización total de 2 es menor al de esta.

Por lo que esta no es la mejor solución. Una vez comprobadas todas las ramas del árbol, sabemos que la mejor solución final es la indicada en la figura 5.11e, y escribe los datos de la misma en un fichero de texto, en el cual están indicados adicionalmente la topología de red a la que corresponde esta ejecución, el número de la asignación de los nodos, la tecnología de enlace empleada, y el criterio en el cual se ha basado el algoritmo para seleccionar esa solución como la mejor.

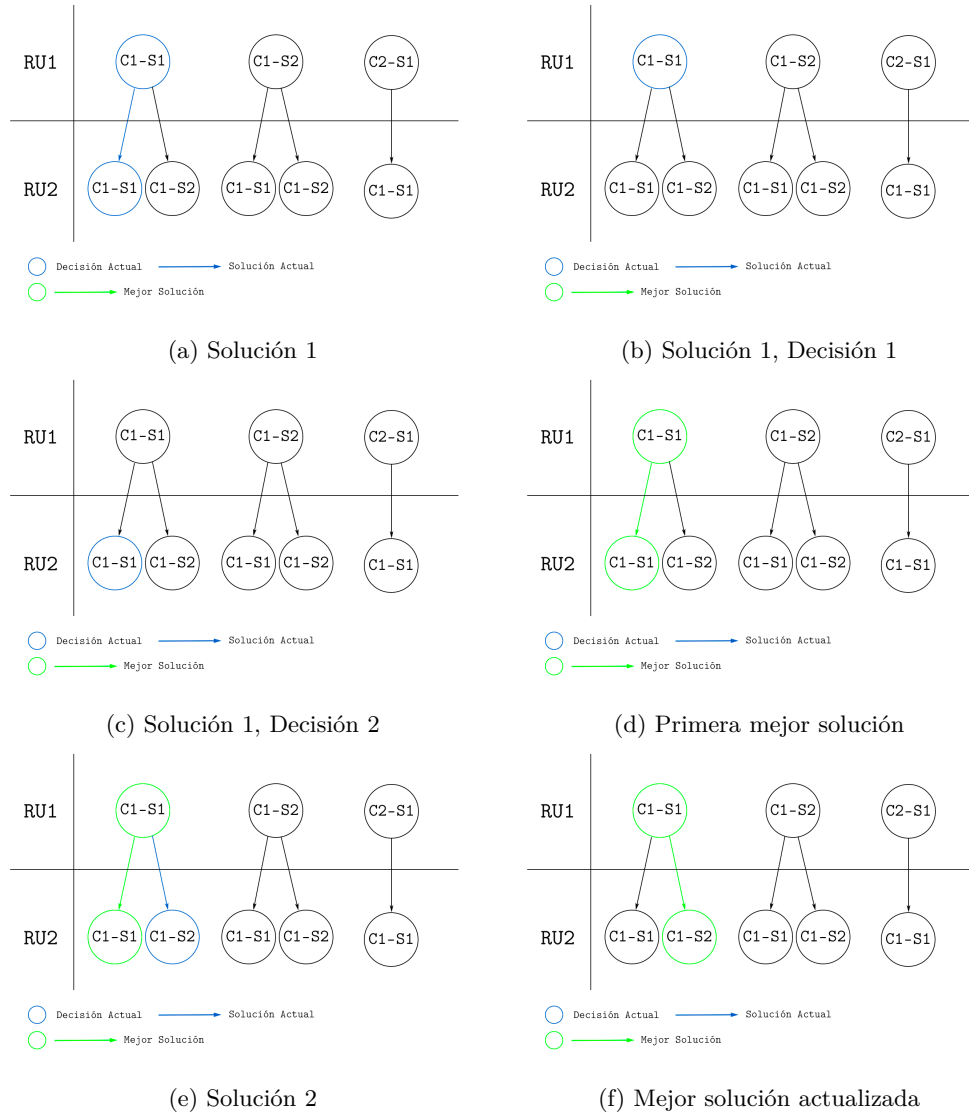


Figura 5.10: Algoritmo Mejor Solucion

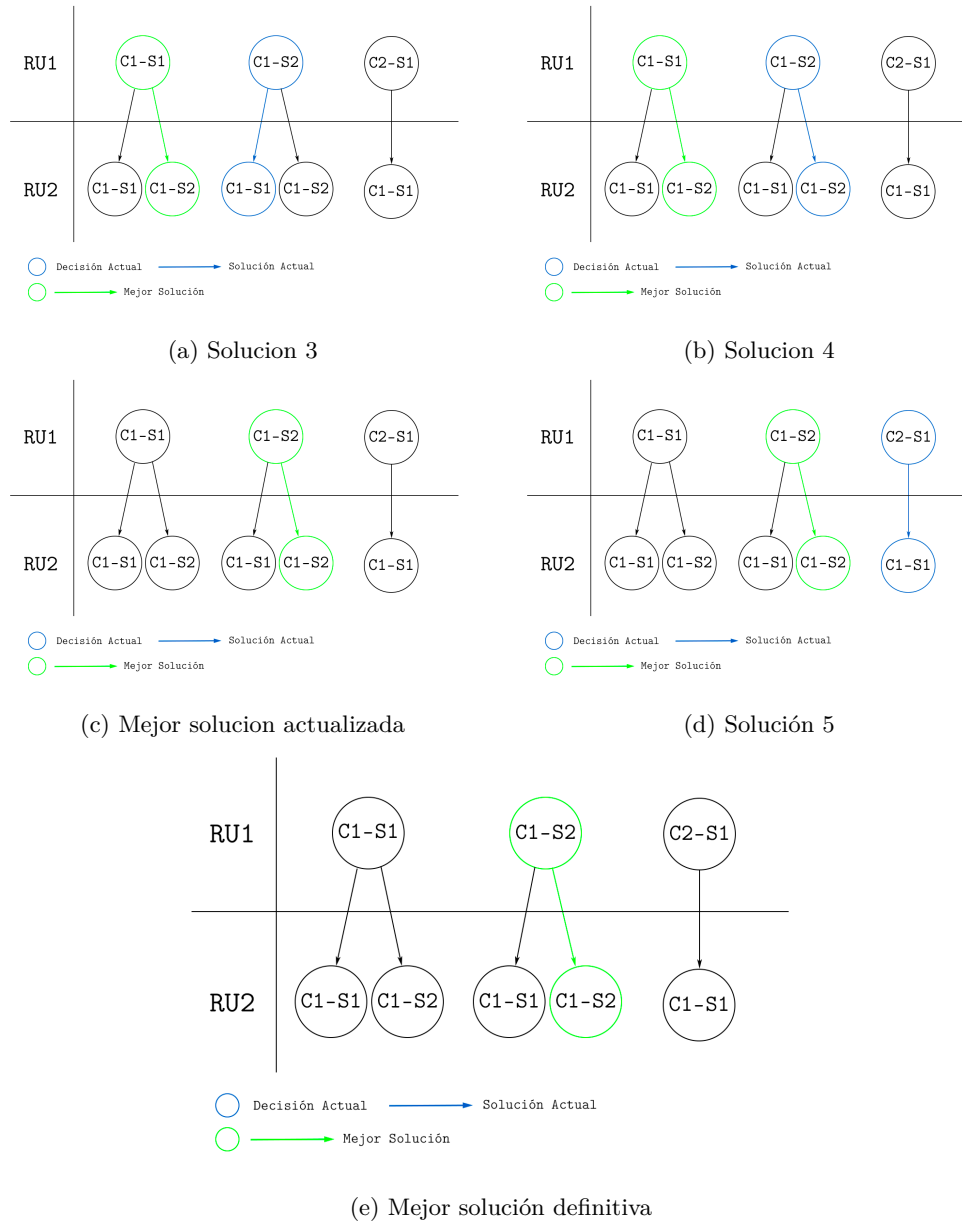


Figura 5.11: Algoritmo Mejor Solucion

Capítulo 6

Datos y Resultados

6.1. Datos con los que se ejecuta el programa

Una vez explicado el funcionamiento del programa y de sus distintos algoritmos con todo detalle, empleando unas redes de ejemplo bastante simplificadas, además de unos datos más simples, lo siguiente es conocer cuales son los datos reales que hemos empleado durante el análisis real de los datos en la ejecución del programa, pues será a partir de estos como obtendremos los resultados que se mostrarán en las siguientes secciones del capítulo.

Es importante aclarar que tanto los niveles de split como las tecnologías de enlace empleadas durante la ejecución del programa son perfectamente configurables, pero vamos a hablar de las que hemos empleado durante esta configuración.

6.1.1. Tecnologías de enlace

Como se explicó en el capítulo de la implementación, y la creación de la red que se va a analizar en cada ejecución del programa, la herramienta realiza un total de 3000 análisis sobre cada una de las topologías de red, repartíendose a su vez entre tres tecnologías de enlace distintas. De esta forma, por cada topología de red se ejecutan un total de 1000 ejecuciones con enlaces de fibra óptica, mil con enlaces de microWave, y por último otras 1000 con enlaces de mmWave.

Obviamente, la tecnología de enlace empleada en cada caso, hará que los resultados obtenidos sean distintos según las características de estas tecnologías, pues son estas las que determinan la capacidad de los enlaces, y el retardo que estos provocan sobre el envío de los datos por los mismos, y son estos parámetros los que condicionan el nivel de centralización que se puede asignar a las distintas estaciones de la red, y por tanto, los que va a tener en cuenta el programa al escoger los caminos y soluciones válidos, de igual forma que la mejor solución entre las posibles.

En este sentido, hay que tener en cuenta que en las tres tecnologías de enlace empleada, la capacidad de los enlaces es siempre la misma, es decir, por ejemplo, en el caso de los enlaces de microWave, la capacidad de sus enlaces es siempre de 2 Gb/s o 2.000 Mb/s, mientras que el retardo provocado por los enlaces estará determinado por la longitud de los mismos, cuyo rango también estará determinado por la tecnología empleada, y la velocidad de propagación, que de igual forma será distinta en distintas tecnologías.

Concretamente, las tecnologías de enlace utilizadas durante la ejecución del programa y sus características son las mostradas en la siguiente tabla:

Tecnología empleada	Capacidad enlaces (Gb/s)	Retardo propagación (μ s)	Distancia (km)
Fibra de 1310 nm	10	1-200	0.2-40
μ Wave	2	1-100	0.3-30
mmWave	8	1-20	0.3-6

6.1.2. Niveles de split

Por otro lado, en la explicación de los algoritmos árbol de soluciones y encontrar solución, empleamos unos niveles de split muy simplificados, con el objetivo de facilitar a su vez la explicación del funcionamiento de ambos algoritmos. En realidad, el programa escoge entre un mayor número de niveles de split distintos. En total, hay 10 niveles de split diferenciados que podrían ser asignados a las estaciones de radio, cada uno de las cuales engloba un cierto número de funciones a centralizar, y con exigencias de delay y capacidad distintas. Estos niveles y sus datos se muestran en la siguiente tabla:

Nivel de split	Funciones centralizadas	Capacidad necesaria (Mb/s)	Retardo máximo exigido (μ s)
A	RRC-PDCP	151	30000
B	PDCP - RLC	151	30000
C	RLC - MAC	151	6000
D	MAC I - MAC II	151	6000
E	MAC - PHY	152	250
F	PHY split I	173	250
G	PHY split II	933	250
H	PHY split III	1075	250
I	PHY split IIIb	1966	250
J	PHY split IV	2457.6	250

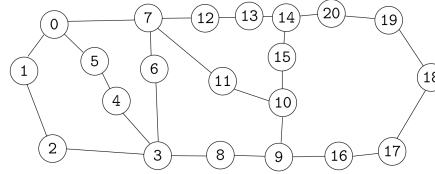
Las exigencias de capacidad y delay de varios de estos niveles de split son muy similares, o incluso iguales entre sí. Por este motivo, se han agrupado en grupos con necesidades similares, de forma que durante la ejecución del programa, este solo comprueba y asigna 5 de estos niveles de split, concretamente los niveles B, D, F, H y J.

Entendiendo que cuando asignamos un nivel de split estamos centralizando tanto las funciones correspondientes a este, como las correspondientes a los niveles inferiores, al asignar un nivel de split a una estación de radio, estamos asignando todos los niveles inferiores al mismo tiempo, y es esto lo que permite comprobar un menor número de niveles de split, sobre todo con capacidades similares, de forma que el hecho de tener un menor número de niveles de split posibles, permite reducir el número de ramas y el tamaño del árbol de soluciones de forma exponencial, lo que causa que el algoritmo sea capaz de encontrar la solución en mucho menos tiempo, y de esta forma permite que se pueda ejecutar el programa de análisis un mayor número de veces en el mismo tiempo. En cualquier caso, los datos correspondientes a los 10 posibles niveles de split están guardados en el programa, por lo que es fácilmente configurable que niveles de split queremos que pruebe y cuales no.

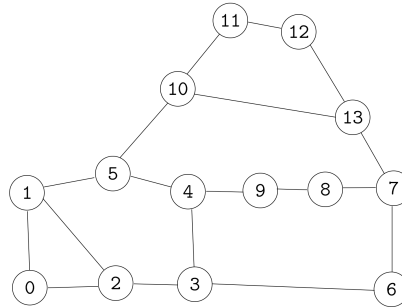
6.1.3. Topologías de red

Por otro lado, hay que tener en cuenta que los resultados que se mostrarán en las siguientes secciones se corresponden a tres topologías de red distintas. Las cuales tienen un diferente número

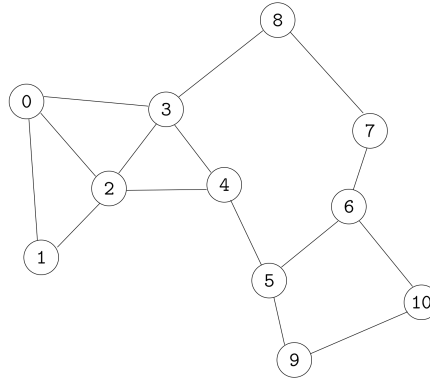
de estaciones de radio, nodos intermedios y de la nube, en función del tamaño que tiene cada una de ella. Estas topologías de red son las mostradas en las imágenes de la figura 6.1



(a) Topología de red 1



(b) Topología de red 2



(c) Topología de red 3

Figura 6.1: Topologías de red

Hay que tener en cuenta que estas topologías únicamente indican que nodos se encuentran conectados entre sí, pues y aunque en el dibujo se encuentren situados a distintas distancias, estas no son representativas a la realidad, pues la longitud de los enlaces (que a su vez determina el retardo provocado por estos según la tecnología empleada en cada caso), se determina posteriormente de forma aleatoria en la función encargada de crear la red. Variando además entre cada ejecución de la herramienta en una sola ejecución del programa. Precisamente, como se explicó en la implementación del programa, por cada una de las 10 asignaciones de los nodos en una topología, se ejecuta la herramienta un total de 300 veces, correspondientes a variar la longitud de los enlaces un total de 100 veces de forma aleatoria sobre tres tecnologías de enlace distintas.

De igual formas, en estas topologías no están indicado el tipo del que es cada nodo, pues la acción del programa encargada de crear la red, determinará sus tipos de forma aleatoria, en

función de los datos de un fichero de configuración, en el que se indican el número de nodos que son de cada tipo, con el objetivo de poder analizar varias redes distintas sobre una única topología, permitiendo de este modo obtener resultados en condiciones distintas, sin tener que introducir manualmente al programa tantas configuraciones de red, sino que al realizar 10 asignaciones de los nodos sobre una misma topología, se obtienen resultados más variados ahorrando bastante tiempo. En este caso, el número de estaciones de radio y de nodos en la nube para cada una de las topologías, son los indicados en la siguiente tabla, mientras que los nodos restantes serán nodos intermedios de la red.

Topología	Nodos totales	Estaciones de radio	Nodos de la nube
1	21	3	3
2	14	4	2
3	11	3	2

Con respecto a la ejecución de la herramienta de análisis sobre estas topologías hay que tener en cuenta, como ya se explicó en la explicación del algoritmo que genera el árbol de soluciones, el tamaño de este árbol crece de forma exponencial, tanto con el número de estaciones de radio, como con el número de caminos que salen de estas, siendo que con la red de ejemplo inicial, que tenía únicamente 3 estaciones de radio, y 4 caminos que partieran de cada una de estas, ya nos encontrábamos con un árbol que podía tener varios miles de soluciones distintas.

Debido a este crecimiento exponencial, aun manteniendo un número de estaciones de radio cercano en estas topologías de red a analizar, el mayor número de enlaces de estas, genera que a su vez el número de caminos que salen de cada estación de radio a un nodo de la nube aumente, lo que genera que el número de posibles soluciones totales pueda ser aún mayor, y si a esto se le suma que sobre cada topología de red se ejecuta el análisis un total de 3000 veces, el tiempo necesario para realizar el análisis y obtener los datos se vuelve terriblemente grande, aun teniendo a mi disposición una capacidad de cómputo bastante razonable. Por este motivo, para que fuera posible realizar la ejecución del algoritmo en un orden temporal inferior a los días, se ha determinado una limitación del número de caminos que parten de cada estación de radio que se van a utilizar en el momento de crear el árbol de soluciones.

Para hacerlo, el algoritmo Paths encargado de encontrar los caminos que van desde cada una de las estaciones de radio hasta los nodos de la nube, incluye una función que ordena los caminos que parten de cada una de estas estaciones de radio, en función de su delay total, ya que al emplear una única tecnología de enlace en cada red a analizar, todos los caminos tienen la misma capacidad. Una vez hecho esto, el algoritmo guarda en la entidad de cada estación de radio, el número correspondiente de caminos en cada caso, escogiendo entre todos los que menor delay total provoquen en el envío de los datos.

Estas limitaciones no son siempre las mismas, sino que dependen de la topología de red en la que nos encontremos, y también de la tecnología de enlace empleada en cada caso. Por ejemplo, los enlaces de mmWave permiten asignar mejores niveles de split, al tener menor delay y más capacidad, por lo que el árbol generado con esa tecnología de enlace tiene más ramas válidas, y por tanto, lleva más tiempo generar y buscar la mejor solución en él, que con otras tecnologías con peores capacidades. Por este motivo, la limitación en el número de enlaces a la hora de utilizar esta tecnología es más estricta, para compensar el aumento en el tamaño del árbol provocado por esas mejores capacidades.

En el caso de la topología 3, que es que tiene el menor nodos y enlaces, y por tanto, la menor complejidad del árbol, solo hemos limitado el número de caminos por cada estación de radio en las ejecuciones con enlaces de mmWave, mientras que en las otras dos se han limitado el número de caminos para las tres tecnologías de enlace. El número de caminos que se han utilizado para cada topología y tecnología de enlace es el mostrado en la siguiente tabla.

Topología	Tecnología enlace	Nº Caminos
1	fibra	5
	μ Wave	5
	mmWave	5
2	fibra	5
	μ Wave	5
	mmWave	4
1	fibra	S.L.
	μ Wave	S.L.
	mmWave	5

6.2. Resultados de la ejecución

Una vez explicadas las distintas topologías empleadas para generar las redes sobre las cuales se va a ejecutar la herramienta de análisis, así como las distintas tecnologías de enlace, y los niveles de split que el programa puede asignar a cada una de las estaciones de radio, además del funcionamiento de los distintos algoritmos que conforman el programa, la última sección de este capítulo consiste en explicar los resultados obtenidos.

Para ello, lo primero es explicar cuales son las métricas que el programa extrae de cada resultado del análisis, es decir, qué datos de la solución seleccionada por el algoritmo pueden ser interesantes para nuestro algoritmo, además del tipo de gráficas que vamos a emplear para mostrar los distintos datos, para finalmente mostrar las métricas extraídas realmente para cada una de las redes en estas gráficas, y explicar las diferencias entre ellas en función de la topología de red a la que correspondan, la tecnología de enlace empleada en cada caso, y el criterio empleado para seleccionar esa solución.

Con respecto a las métricas extraídas, únicamente hemos extraído 4 valores distintos de cada una de las redes sobre las que se ha ejecutado el algoritmo, que son precisamente las 4 métricas empleadas por los distintos criterios de selección, que emplea el algoritmo encargado de escoger la mejor solución en el árbol de soluciones. Tiene sentido que las métricas que determinan que solución es mejor, sean a su vez las más relevantes a la hora de analizar las distintas soluciones escogidas. En este caso, cada vez que el programa finaliza la ejecución del algoritmo, escoge la mejor solución y encuentra con ello la mejor solución del árbol en función de uno de los criterios, guarda los valores del delay máximo, el delay total, el split mínimo y el split total correspondientes a esa solución en específico, en un archivo de texto en el cuyo nombre se indican tanto el criterio empleado en ese caso, como la topología de red a la que corresponde y el número de la organización de los nodos en el que nos encontramos dentro de la topología.

Una vez el programa ha ejecutado la herramienta de análisis sobre todas las redes explicadas, y ha guardado todas las métricas de la mejor solución para todas las topologías de red, todas las organizaciones, todas las tecnologías de enlace, y todos los criterios en sus correspondientes archivos, el siguiente paso es analizarlos. Para ello, al tener un total de 36.000 soluciones, que se corresponden a las 9.000 fotografías analizadas multiplicadas por los 4 criterios empleados para escoger la mejor solución en cada una de ellas. Con este número de soluciones, es fácil de ver que es imposible comparar cada una de estas soluciones por separado, y además, la idea de hacer un análisis tan amplio es precisamente tener una visión de conjunto, de forma que podamos observar como afectan los cambios en los parámetros del análisis al resultado obtenido.

Por este motivo, el primer paso es agrupar las soluciones que se correspondan a una misma topología de red, tecnología de enlace, y criterio de selección, de forma que tendremos un total 36 grupos (3 topologías * 3 tecnologías de enlace * 4 criterios para escoger solución), cada uno de los cuales tendrá 1000 soluciones, que analizaremos en conjunto. De esta forma, lo que se comparará es la distribución de las distintas métricas en cada uno de estos grupos, en relación con los demás,

de forma que podamos de que forma cambian estos valores en función de la topología en la que nos encontremos, las tecnología de enlace empleada en cada caso, o el criterio empleado para escoger la solución sobre una misma red.

Para poder analizar estas diferencias, lo que hemos hecho es dibujar mediante Matlab los valores de las métricas de cada uno de estos grupos, empleando tipos de gráficas distintos para las métricas que muestran el delay y para las que muestran el nivel de centralización, debido a la diferencia entre la variabilidad de los datos en cada tipo de métrica.

6.2.1. Delay

En el caso del delay, se dibuja en una gráfica llamada Box Plot, de la cual un ejemplo es la mostrada en la figura 6.2. Este tipo de gráfica es mejor para mostrar los resultados del delay, puesto que estos son muy variables, y el box plot permite mostrar gráficamente el rango en el que se mueven los valores. Siendo que el cuadro azul es donde se encuentran la mayoría de los valores, concretamente todos los valores entre el 25 y el 75 %, que la línea roja dentro de las cajas indica la mediana, o el 50 % de los datos, y las líneas negras indican los valores extremos, es decir, en principio, los valores mínimo y máximo de todos los datos.

Digo en principio, porque es así excepto en el caso de que existan datos atípicos en la distribución. Estos datos atípicos son aquellos que se alejan de los bordes de la caja más de 1,5 veces la longitud de la misma. Es decir, que las líneas negras, estarán como muy lejos a 1,5 veces la longitud de la caja azul, y los valores que queden por encima o por debajo de la línea superior e inferior respectivamente, se marcarán individualmente como valores raros o atípicos, empleando una cruz roja.

Explicando esto con el primero de los 12 box plot que se muestran en la figura 6.2, que se corresponde con los datos de delay máximo de todas las soluciones encontradas según el criterio de mínimo delay sobre la topología 1 y empleando la fibra óptica como tecnología de enlace, la caja azul va desde los 220 ms, lo que nos indicaría que únicamente el 25 % de los datos estaría por debajo de ese valor, hasta los 400 ms, por lo que el 75 % de los datos estaría por debajo de este. Esto nos dejaría con que la longitud de la caja es de 180. Con eso, podemos saber, que las líneas negras podrían encontrarse como mucho en los valores de $220 - 1.5 * 180 = -50$ ms y $400 + 1.5 * 180 = 670$ ms. En el caso del valor de delay más pequeño, se encuentra por encima de -50, por lo que esa línea se dibuja a la altura de ese valor, que en este caso son 30 ms aproximadamente. Mientras que en el caso del valor máximo, hay varios valores que se encuentran por encima de 670 ms, por lo que la línea se dibuja a la altura de 670 ms, y esos valores por encima, se indican como valores atípicos mediante cruces rojas, teniendo en este caso 4 valores atípicos entre los 695 y los 720 ms, y un quinto en 760 ms. Y por último la línea roja que marca la mediana, o el 50 % de los datos, se encuentra en 300 ms.

La ventaja de este tipo de gráfica, es que permite conocer la distribución completa de los datos, mediante una única columna o una única gráfica relativamente sencilla, aun teniendo datos con un nivel de variabilidad relativamente alto, de forma que en apenas 12 columnas se pueden dibujar 12 box plot, que muestran un total de 12.000 resultados de delay total, correspondientes a todas las configuraciones distintas que se analizan sobre una topología completa.

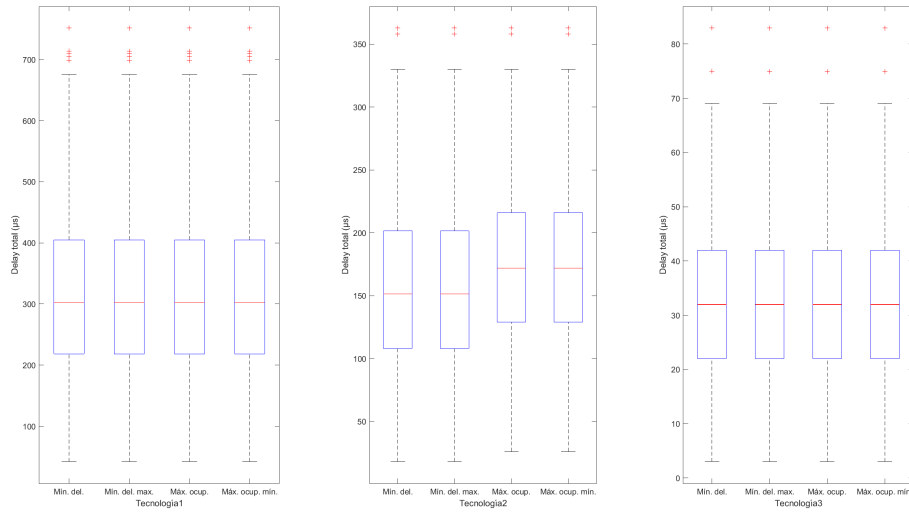


Figura 6.2: Box Plot

6.2.2. Nivel de split

Por otro lado, en el caso de las métricas que muestran el nivel de split de la solución escogida, tanto para el total de todas las estaciones de radio, como para el nivel de split mínimo entre estas, no tiene sentido emplear una gráfica box plot, debido a que los niveles de split que se pueden obtener como resultado para una misma topología y tecnología de enlace siempre van a tener unos valores determinados, debido a que dependen de los niveles de split asignables, del número de estaciones de radio, y de la tecnología de enlace empleada.

Explicado sobre las redes que hemos analizado todas tienen 4 o menos estaciones de radio, a las cuales se les pueden asignar un total de 5 niveles de split distintos, dos de los cuales causan la misma ocupación de la red, por lo que realmente tendríamos que considerar únicamente 4 niveles de split distintos a la hora de calcular esta métrica (puesto que su valor depende precisamente de esa ocupación), lo que nos da que la métrica de mínimo nivel de split solo puede tener 4 valores, que se corresponden con esos 4 niveles de split asignables, mientras que, mediante combinatoria, obtenemos que la métrica del split total en el caso de la topología con el mayor número de estaciones de radio, puede tener un máximo de 35 valores distintos (la combinación de un conjunto de 4 elementos (4 niveles de split), seleccionando 4 elementos (las 4 estaciones de radio), y pudiendo repetir valores, da un total de 35 posibles combinaciones, y con ello 35 posibles valores para esa métrica).

Con estos 35 valores posibles distintos ya no tendría mucho sentido emplear un box plot para representar los datos, y además, resulta que una vez ejecutado el programa, nos encontramos con que en el caso de la topología y tecnología de enlace en la que obtenemos el mayor número de valores distintos para esta métrica, únicamente hemos obtenido 6 valores distintos. Esto se debe a que al emplear una única tecnología de enlaces para todos los enlaces sobre una misma topología, los niveles de split que se asignan a las distintas estaciones de radio de la misma sea varíe muy poco entre las distintas ejecuciones de la herramienta de análisis sobre una misma topología, y lo que provoca que el número de valores que encontramos finalmente se reduzca mucho.

En resumen, la idea es que los valores de esta métrica son muy poco variables, mientras que el objetivo del box plot es precisamente mostrar datos con una alta variabilidad, de forma que

podamos conocer la distribución de estos, como en el caso de las medidas de delay, que pueden variar completamente entre una ejecución y otra del programa, ya que los delays de cada uno de los enlaces son asignados de forma totalmente aleatoria. Por este motivo, el nivel de split, se muestra en un tipo de gráfica que permite mostrar más claramente los valores obtenidos.

Concretamente, se emplean dos tipos de gráficas para mostrar estos valores. Del primer tipo son las gráficas mostradas en la figura 6.3. Estas gráficas muestran el porcentaje de resultados, expresados en un valor del 0 al 1, para los cuales la métrica analizada tiene cada uno de los valores obtenidos, entre las mil ejecuciones que se han realizado para cada una de las topologías de red, tecnología de enlace y criterio de selección de la solución. Como se puede observar, cada figura de este tipo tiene un total de 12 gráficas, que se corresponden a una tecnología de enlace (en horizontal), y criterio (en vertical), y dentro de cada una de estas gráficas hay una barra por cada combinación de niveles de split distinta que hemos encontrado entre esas mil ejecuciones, expresada como la ocupación de la red causada por estos niveles de split. En este caso, la figura 6.3 muestra los valores de split total para todas las ejecuciones sobre la topología 1.

Entrando en detalle, para entender lo que muestra cada una de las gráficas de la figura, si por ejemplo nos fijamos en gráfica superior izquierda de la figura 6.3, que es la que muestra los valores de split total en el caso de enlaces de fibra óptica, y según el criterio de selección de solución de mínimo delay, podemos observar que solo ha habido 4 valores distintos para esta métrica entre las mil soluciones analizadas, y que entre ellas, en aproximadamente el 12 % de las ejecuciones el nivel de split total de la solución escogida es de 453 Mb, en el 20 % es de 2759 Mb, en el 34 % es de 5065 Mb y en el otro 34 % restante, de 7371 Mb. Mientras que en la gráfica que muestra las soluciones obtenidas por el mismo criterio, pero empleando en este caso enlaces de mmWave (que sería la superior derecha), en todas las soluciones seleccionadas por el programa, el nivel de centralización total es de 7271 Mb, lo que quiere decir que en todas las ejecuciones se ha asignado el nivel de split máximo a las 3 estaciones de radio de la topología.

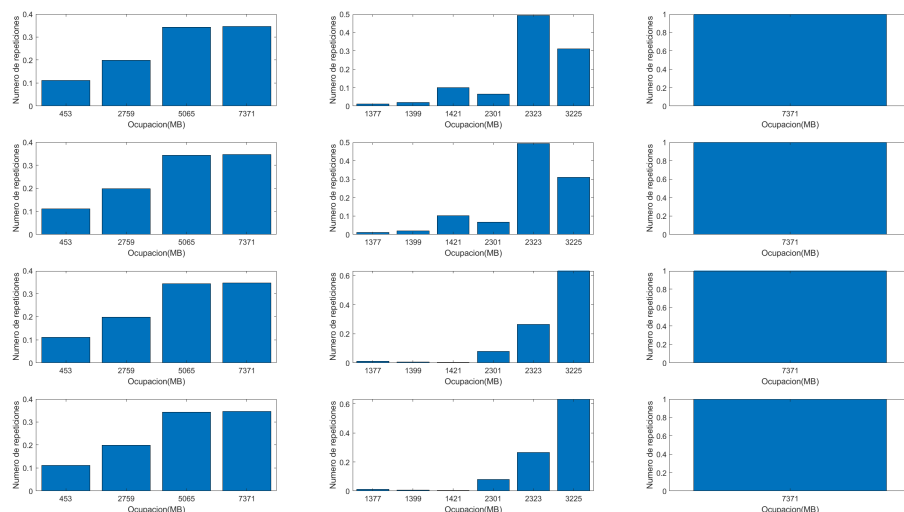


Figura 6.3: Histograma

El segundo tipo de gráfica, que es el mostrado en la figura 6.4, a diferencia del primero, busca mostrar los valores de nivel de split de forma más sencilla y rápida. Para ello, simplemente muestra el valor medio de esa métrica para las mil soluciones para una tecnología de enlace y criterio. Mostrando en una columna cada uno de estos valores. Este tipo de gráfica tiene la ventaja de ser más fácil de representar, al necesitar una única columna (y un único valor) por cada grupo

de 1000 soluciones, y de igual forma, permite observar más fácilmente la diferencia general del nivel de centralización obtenido en función de la tecnología de enlace y del criterio empleado para seleccionar la solución para una misma topología. Pero a cambio, tiene la desventaja de que no se puede observar la distribución de los valores de este nivel de centralización entre todas las soluciones.

Como se puede observar, en una figura de este tipo, hay 12 columnas que se corresponden a la multiplicación de las tres tecnologías de enlace y los cuatro criterios de selección de solución, igual que en el caso de los box plot, con la diferencia de que en ese caso esa columna mostraba la distribución de todos los datos, mientras que en este solo busca comparar los resultados de cada caso entre sí, dando un único valor que muestra la media de todos los resultados para cada uno de estos.

Sabiendo los dos tipos de gráficas que se pueden utilizar para mostrar el nivel de split asignado a las distintas estaciones de red en cada caso, en el análisis de los resultados se utilizarán uno u otro, en función de si se quiere analizar de forma más precisa la forma en la que se distribuyen los resultados para todas las soluciones de un mismo caso, o si el objetivo es simplemente comparar los resultados de distintas situaciones entre sí de forma más rápida.

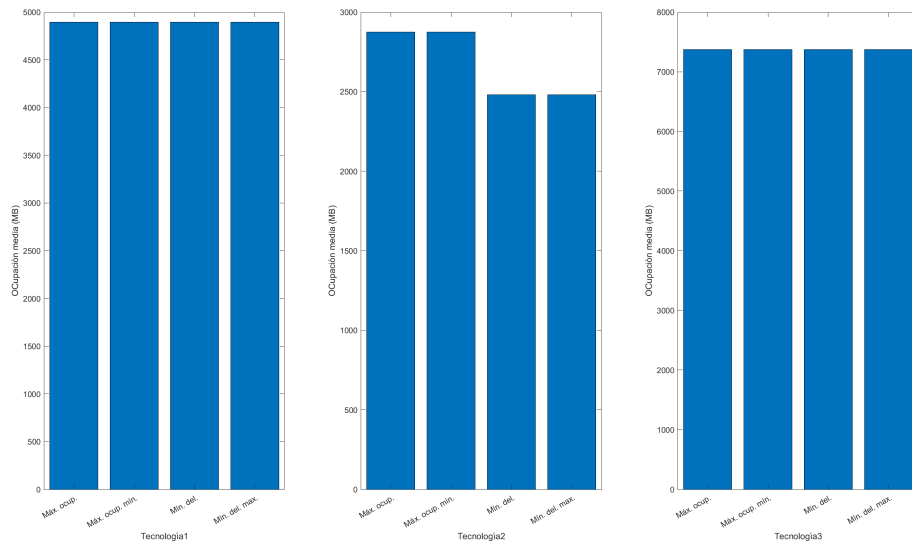


Figura 6.4: Media

6.3. Análisis de los Resultados

Después de explicar en la anterior sección los distintos tipos de métricas que hemos extraído de la ejecución de la herramienta sobre las redes analizadas, así como los distintos tipos de gráficas empleados para mostrar estos datos en función del tipo de cada uno, en esta sección se verán estas gráficas, y se analizará de qué forma las métricas extraídas dependen de los distintos factores que afectan a la red. Es decir, que se estudiará y explicará, mostrando las distintas gráficas dibujadas, en qué medida varían cada una de las métricas analizadas en función del tamaño de la red, o de la tecnología de enlace empleada. Para hacer esto, vamos a separar el análisis los resultados en dos apartados, uno en el que analizaremos los datos del delay de los caminos por los que se envían las funciones de red en las soluciones seleccionadas, mostrados en gráficas del tipo box plot, y otro en el que se analizará el nivel de split, mediante la ocupación de la red provocada por el split

asignado a las estaciones de radio de cada topología en las distintas soluciones.

A la hora de analizar los resultados de las distintas gráficas, es muy importante tener en cuenta los escenarios sobre los que se ha ejecutado el programa, pues tanto la ocupación, como el delay de los caminos asignados, van a variar en función de parámetros como el tamaño de la red, el retardo asociado a los enlaces, y el número de estaciones de radio a las que asignar un nivel de split. Es por esto, que los datos mostrados en la primera sección de este capítulo cobran especial relevancia, y van a ser muy tenidos en cuenta durante el análisis de las distintas métricas realizado en esta sección.

6.3.1. Delay

Lo primero que vamos a analizar es como varía el delay, en función de la topología de red, la tecnología de enlace, y el criterio empleado en cada caso. Para esto, vamos a analizar las métricas de delay total, y delay máximo, que nos muestran los resultados de toda la red en conjunto, y de la estación de radio con el camino en peores condiciones de delay, respectivamente.

Como se explicó en la anterior sección, para cada una de estas métricas nos encontramos con 3 figuras, una por cada topología de red, cada una de las cuales muestra un total de 12 gráficas que muestran los datos para todas las combinaciones de tecnologías de enlace y criterios.

Delay Total

El delay máximo mide la suma de los delays de los caminos asignados a cada una de las estaciones de radio, por los que se envían las funciones de red correspondientes al nivel de split indicado en la solución escogida por el algoritmo de selección.

Es imposible analizar una a una las 36 gráficas que muestran los datos de delay máximo, que encontramos en las figuras 6.5 y 6.5 y compararlas entre sí, así que el objetivo es extraer una conclusiones generales del conjunto de las gráficas.

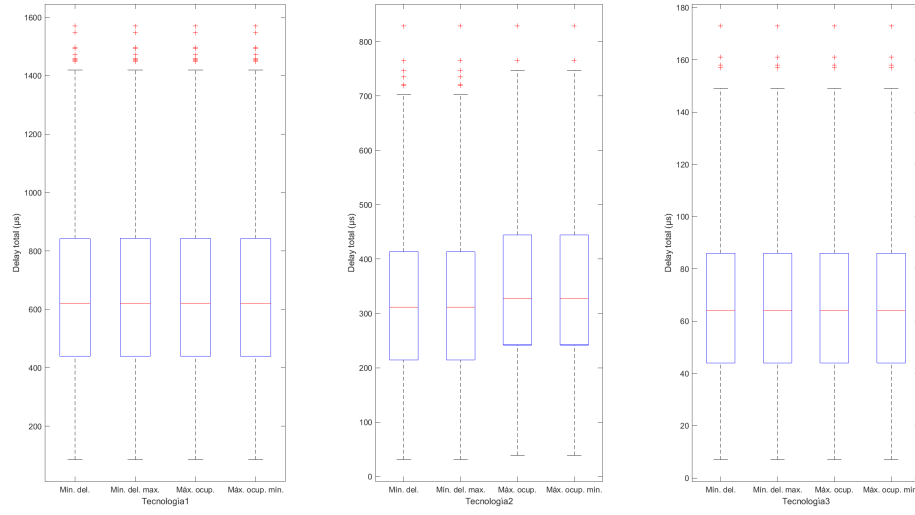
Lo primero que podemos observar es que los resultados para cada topología y tecnología de enlace, coinciden para todos los criterios de selección de solución, excepto en el caso de los enlaces de μ Wave, para los cuales, los criterios basados en la ocupación de los enlaces causan un mayor nivel de delay que los basados en el delay, lo cuál tiene sentido, pues quizá el camino que permite asignar un mayor nivel de split, no tiene por qué ser necesariamente el más corto en todos los casos.

Lo siguiente es observar un grupo de gráficas en concreto, que en este caso serán las de la figura 6.5a, que se corresponden con la primera de nuestras topologías, y comparar los resultados según la tecnología de enlace empleada, pudiendo observar que el retardo total es proporcional al retardo medio esperado para cada una de las tecnologías de enlace empleadas, siendo que los enlaces con meno retardo son los de mmWave, y los de mayor retardo los de fibra óptica. Conclusión que se mantiene en las gráficas de las figuras 6.5b y 6.6a

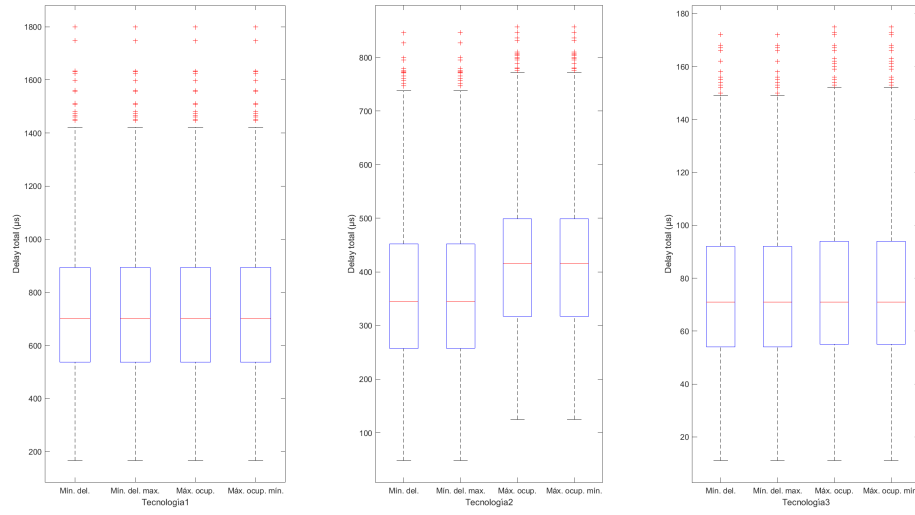
Por último, lo último que podemos comparar son los resultados de las topologías entre sí. Si nos fijamos en todas las gráficas, para una misma tecnología de enlace y criterio, nos encontramos con que la topología 2 es la que muestra mayor retardo total, y la topología 1 la que menos.

Este resultado podría contrastar con lo esperado, pues es la topología 1 la que tiene la red de mayor tamaño, pero esta métrica mide la suma de los retardos de todos los caminos de todas las estaciones de radio de la red, y resulta que la topología 2 es la que tiene más estaciones de red. Sin embargo, si dividimos el retardo total entre el número de RUs, nos encontramos con que la

topología 1 tiene un mayor retardo por RU y camino que la topología 2.

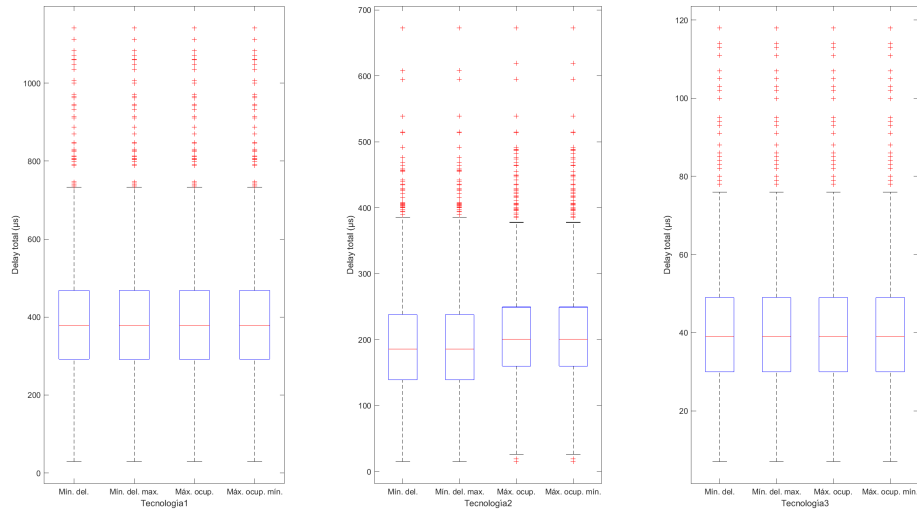


(a) Topología 1



(b) Topología 2

Figura 6.5: Delay Total



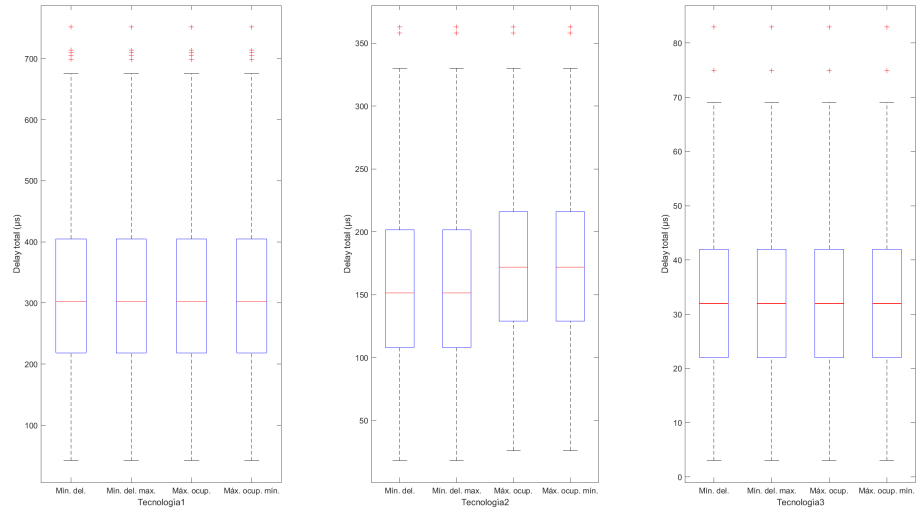
(a) Topología 3

Figura 6.6: Delay Total

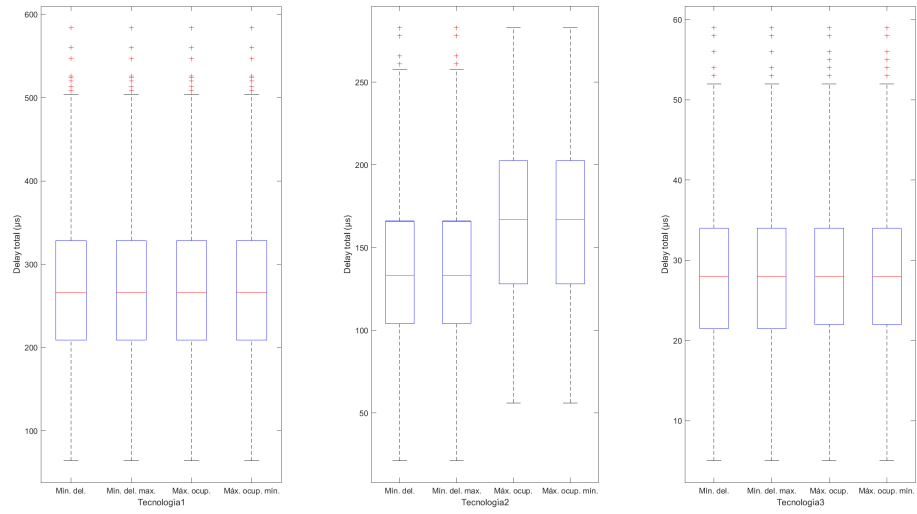
Delay Máximo

El delay máximo es la métrica que guarda los datos del camino con más delay entre los asignados en la solución a cada una de las RUs de la red.

En este caso, las gráficas que muestran esta métrica se encuentran en las figuras 6.7 y 6.8, y los resultados son los mismos que los encontrados en el caso del delay total, con la diferencia de que en este caso, la topología 1 es la que muestra el mayor delay en todos los casos. Esto se debe, a que esta métrica solo analiza el delay de un único camino para una única estación de red, y no la suma de todos ellos, y por este motivo, la red de mayor tamaño es la que muestra el mayor delay máximo.

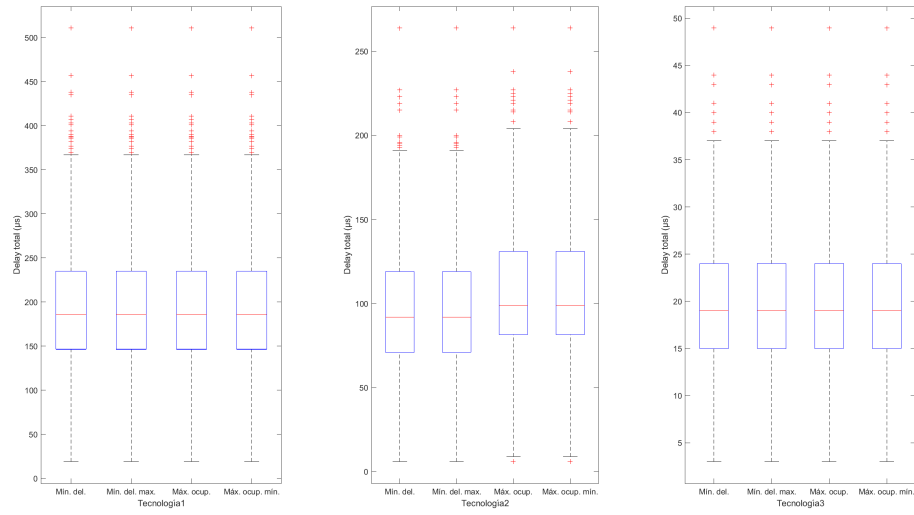


(a) Topología 1



(b) Topología 2

Figura 6.7: Delay Máximo



(a) Topología 3

Figura 6.8: Delay Máximo

Conclusión

La conclusión a la que podemos llegar, al respecto del delay, tanto en el caso del delay total, como del máximo, es que estas métricas aumentan según el tamaño de la red, pues en una red más grande suele haber más saltos entre una RU y su CU asignada, y por tanto el delay del camino que conecta ambos nodos aumenta.

6.3.2. Nivel de Split

La siguiente métrica a analizar es el nivel de split, que en este caso se muestra mediante la ocupación de la red. La idea para entender esto, es que cuanto mayor sea el nivel de split asignado a una estación de radio, más funciones se envían a través de la red para ser centralizadas en la nube, y estas funciones causan una cierta ocupación de la red. Así que en lugar de indicar los niveles de split como tal, la métrica utilizada para saber si el nivel de split es mayor o menos, es la de la cantidad de información que se envía a la nube para centralizar sus funciones. Es decir, a mayor nivel de split, mayor ocupación.

En este apartado se van a analizar tanto el split total acumulado de todas las estaciones de radio de una red, como el split mínimo de alguna de ellas. Utilizando para la primera métrica las gráficas del tipo histograma para mostrar más datos, y en el segundo la media para hacer una comparativa más directa, y de esta forma mostrar también como se pueden extraer distintos datos de las distintas gráficas.

Split Total

Las gráficas de split total, igual que en el caso del delay total de la anterior sección, muestran la suma de todos los niveles de split, o de todas las ocupaciones de todas las estaciones de red de una topología.

En este caso, las gráficas que vamos a analizar, son las mostradas en las figuras 6.9 y 6.10, y lo primero que hay que tener en cuenta es qué muestran cada una de las tablas de esas figuras. Dentro de esas figuras hay 3 subfiguras, cada una de las cuales muestra los datos de una topología, y dentro de cada una de ellas, hay un total de 16 gráficas, las cuales se corresponden de izquierda a derecha a las tecnologías de enlace de fibra, μ Wave y mmWave, y de arriba a abajo a los criterios de delay mínimo, mínimo delay máximo, máxima ocupación, y máxima ocupación mínima.

Observando total las gráficas de las figuras 6.9 y 6.9 en conjunto, podemos llegar a la misma conclusión que en el caso del split total, y es que los criterios de selección de solución llevan a pocas diferencias entre los resultados. Siendo que únicamente encontramos resultados distintos entre los criterios de delay y los de split para la tecnología de enlace 2 de las tres topologías (igual que en el caso del split), y adicionalmente para la tecnología 3 de la topología 2, mientras que el resto de gráficas para una misma topología y tecnología de enlace coinciden.

Lo siguiente en lo que nos podemos fijar, en la comparativa global, es la forma en la que la topología de red y el número de estaciones de radio afectan al nivel de split total. En este caso, igual que en el anterior, nos encontramos con que la topología con el mayor nivel de split total, independientemente de su tecnología de enlace es la Topología 2, pero igual que en el caso del delay, esto se debe a que hay una estación de radio más.

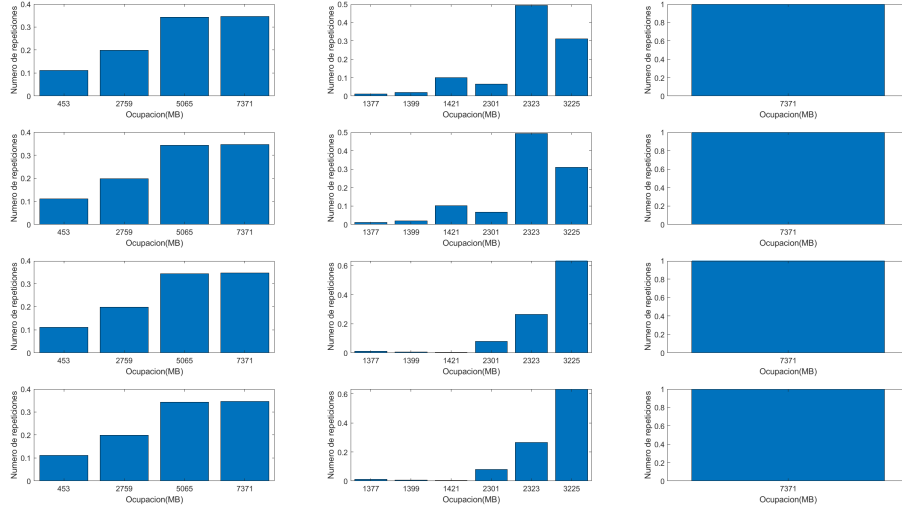
Así que las gráficas que vamos a comparar son las de las figuras 6.9a y 6.10b, que son las que tienen el mismo número de RUs, y nos encontramos con que la topología 3 tiene una mayor nivel de split para las tecnologías de fibra y μ Wave, ya que en el caso de los enlaces de mmWave, el nivel asignado es el máximo posible en todos los casos. Estos resultados se deben a que además de la capacidad de los enlaces, el nivel de split asignable también se encuentra limitado por el delay de los caminos. Y por este motivo, si los caminos son más largos, debido al mayor tamaño de la red, hay veces en las que aunque un camino tenga una capacidad que permita asignarle un cierto nivel de split, el retraso acumulado de todos sus saltos, haga que ya no se le pueda asignar ese split, mientras que en el caso de los enlaces de mmWave, el retardo de cada salto es tan pequeño, que en redes del tamaño de las nuestras, nunca nos encontramos con ese problema.

Por último, el último paso es analizar una topología en concreto, para analizar la influencia de

la tecnología de enlace en los resultados obtenidos para una misma red, siendo que en este caso vamos a analizar las gráficas de la figura 6.10b.

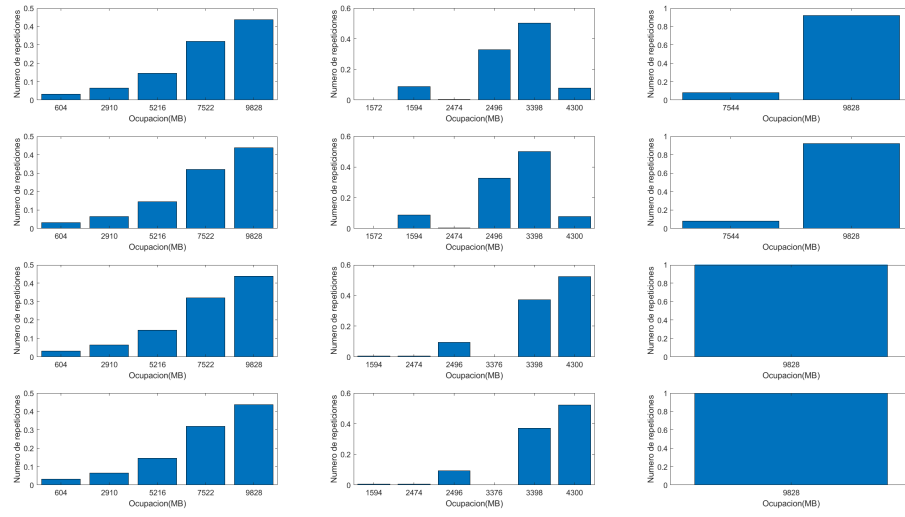
Como podemos observar, utilizando enlaces de mmWave, podemos asignar el máximo nivel de split en todos los casos, mientras que en los otros dos casos la variedad de niveles de split asignados es más amplia, encontrando que en las redes con enlaces de μ Wave, el nivel de split asignado se encuentra limitado principalmente por la capacidad de sus enlaces, que es de 2GB, siendo que si queremos enviar el split de dos RUs distintas por el mismo enlace, el nivel de split asignable a la segunda tendrá que disminuir, y observando también que el máximo nivel de split que se puede asignar a una RU es el H, ya que el nivel de split J, que sí que se puede enviar por las otras dos tecnologías, exige más capacidad que lo que permiten los enlaces de μ Wave.

Pero en cualquier caso, las gráficas realmente interesantes, son las que se corresponden a las redes en las que empleamos fibra óptica, pues en la mayoría de estos el nivel de split asignado es el máximo para todas las RUs, pero sin embargo, hay un pequeño porcentaje de casos en los que se asigna el nivel de split mínimo a todas las RUs. Esto se debe a que los enlaces de fibra tienen la capacidad más alta entre las tecnologías empleadas, por lo que por norma general permiten que se asignen niveles de split altos, pero al mismo tiempo añaden el mayor retardo entre las tecnologías empleadas, por lo que si nos encontramos con un caso en el que los enlaces sean especialmente altos, el nivel de split se verá muy limitado por el delay de los caminos, aunque estos tengan la capacidad de enviar muchos más datos.

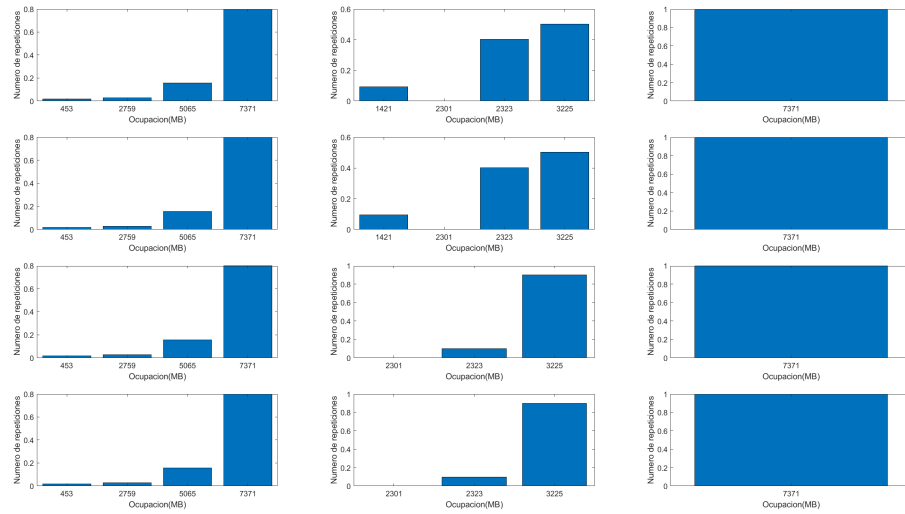


(a) Topología 1

Figura 6.9: Split Total Histograma



(a) Topología 2



(b) Topología 3

Figura 6.10: Split Total Histograma

Split Mínimo

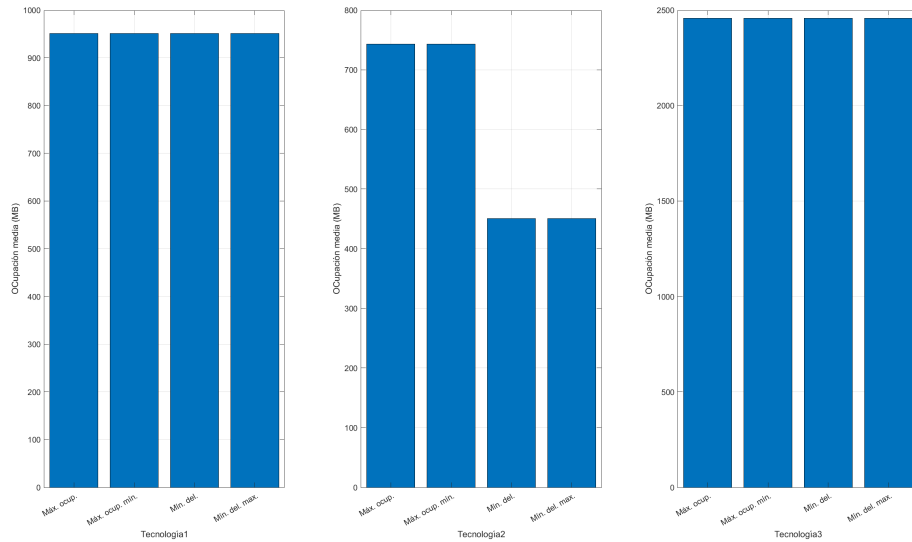
El split mínimo mide cuál es el menor nivel de split asignado a alguna de las estaciones de radio de la red. En este caso, las gráficas que vamos a analizar son las mostradas en las figuras 6.11 y 6.12, siendo a diferencia del caso anterior gráficos que muestran la media, para darnos una visión general más simplificada de los mismos.

Lo primero que observamos, es que en este caso, igual que en el anterior, la variación de los datos según los criterios de selección de la solución es mínima, salvo en el caso de los enlaces de μ Wave, esto se debe a que al tener estos una capacidad muy pequeña, si el criterio de selección de

solución es el de delay mínimo, siempre que haya dos RUs que compartan el mismo enlace por ser el de menor retardo, el nivel de split asignable a uno de ellos bajará mucho, por lo que a su vez bajará el nivel de split mínimo. Mientras que si nos centramos en seleccionar las soluciones con mayores niveles de split, se buscará que los enlaces no se compartan, y el nivel de split mínimo subirá.

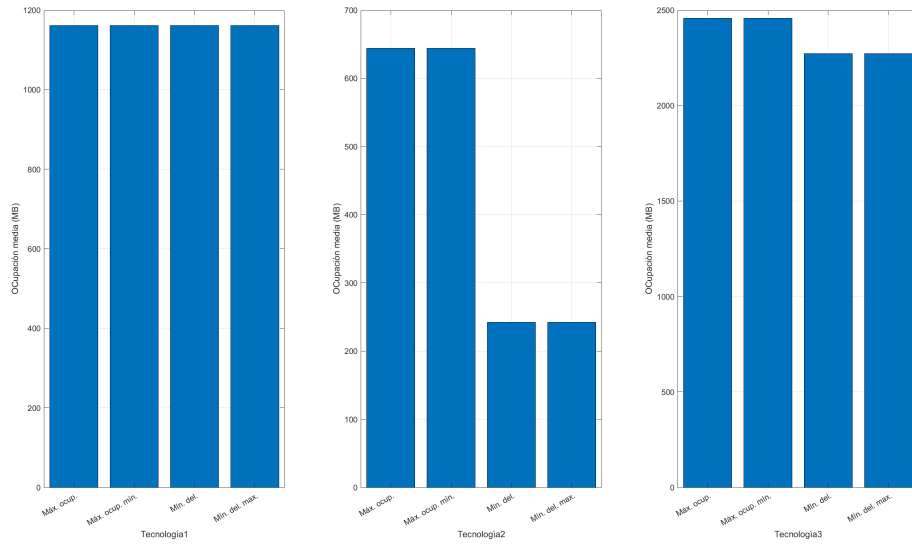
Lo siguiente a tener en cuenta es como la topología de red afecta a los resultados obtenidos, encontrando que topología 1, que es la red más pequeña, es la que tiene los mayores niveles de split mínimos, mientras que es la topología 2 la que tiene los menores niveles de split mínimo más bajos en algunos casos, cuando lo esperable es que fuera la topología 1, que es la red más grande. Esto puede deberse a que la topología 2 tiene una estación de radio más, lo que causa que haya un camino más que seleccionar, que puede coincidir, y causar que a alguna de las RUs de la red haya que asignarle un nivel de split mucho más bajo.

Por último, nos encontramos con que la tecnología de enlace que tiene el mayor nivel de split mínimo medio para todos los criterios es mmWave, mientras que la que tiene el más bajo es μ Wave, por el mismo motivo que se ha explicado en los casos anteriores, que es que en esta tecnología hay muchos casos en el que alguna de las RUs tiene que tener un nivel de split bastante bajo, debido a la capacidad limitada de los enlaces.

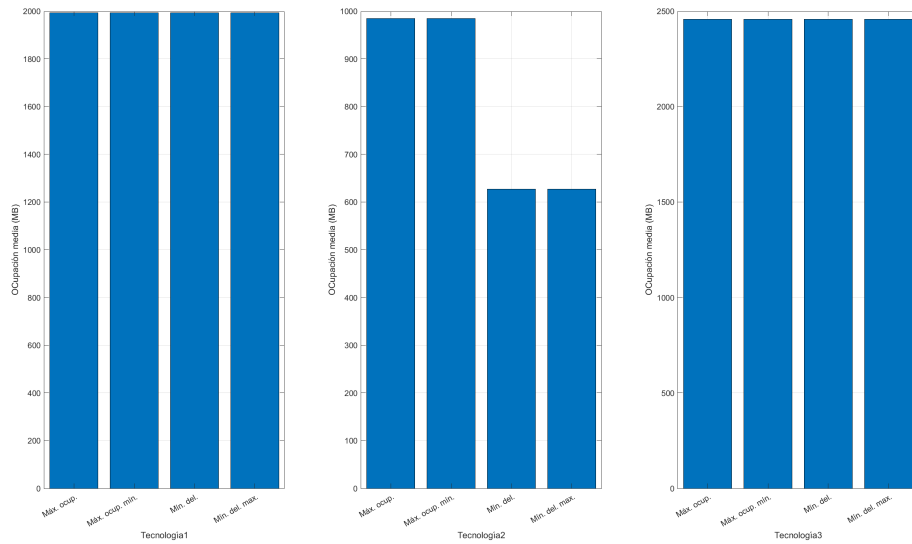


(a) Topología 1

Figura 6.11: Split Minimo Media



(a) Topología 2



(b) Topología 3

Figura 6.12: Split Mínimo Media

Conclusión

Las conclusiones a las que podemos llegar en esta sección son relativamente variadas, pues dependen de varios factores del escenario en el que nos encontremos y de la métrica que estemos analizando, pero la idea general es que lo que más limita el nivel de split que podemos asignar es la capacidad de los enlaces, aunque el delay de los mismos también es relevante a este respecto, por lo que una red más grande, es decir, con más delay en sus caminos, tenderá a disminuir los niveles de split que se le pueden asignar, al haber más probabilidades de no cumplir con los criterios de delay, o por asignar enlaces comunes a varios caminos de distintas RUs.

Por el otro lado, la otra conclusión es que los resultados de split mínimo para una situación tienen cierta concordancia con lo del split total para la misma, pero, que una cierta topología tenga un split mínimo muy bajo, no tiene que implicar necesariamente que su split total lo sea, pues puede ser porque hay una única RU a la que no se le ha podido asignar un split mayor.

6.3.3. Criterios de selección de solución

Además de las conclusiones particulares para ambos tipos de métricas, de este análisis se puede extraer la conclusión de que los criterios de selección de solución solo afectan a algunos de los resultados obtenidos, encontrándonos que en el caso de las métricas del delay, para una misma topología y tecnología de enlace solo hay unas pequeñas diferencias entre los resultados empleando criterios que priorizan un menor delay, y los que priorizan la mayor ocupación, y que tanto para las métricas de delay, como para las de nivel de split, y se obtienen exactamente los mismos resultados entre los criterios de mínimo delay y mínimo delay máximo, así como entre la máxima ocupación y la máxima ocupación mínima.

Esto se debe a varios motivos, el primero, que normalmente la solución con la máxima ocupación total, tiende a tener a su vez la máxima ocupación mínima, y sucede lo mismo en el caso del mínimo delay y el mínimo delay mínimo, y además, los criterios de selección funcionan de forma combinada, es decir, entre dos soluciones con misma ocupación total, el programa escoge la que tiene la máxima ocupación mínima. Además, las redes analizadas son de un tamaño pequeño, y con unas condiciones muy homogéneas, lo que causa que los resultados para casi todos los análisis sean muy similares, aun habiendo realizado un número de estos bastante grande, lo que implica que no haya cientos de caminos entre los que escoger que nos lleven a distintas soluciones según el criterio escogido.

Y la conclusión a la que se puede llegar a partir de esto, es que podría tener sentido, y ser interesante, realizar un análisis sobre redes de mayor tamaño, con condiciones y tecnologías menos heterogéneas, y más similares a las que nos podemos encontrar en el mundo real, utilizando además criterios de selección de soluciones menos genéricos, y más adaptados a las necesidades particulares de cada red, puesto que de este modo, los resultados obtenidos serían del mismo modo más variados, pero que aún así, se ha podido comprobar el sentido que tiene utilizar distintos criterios de selección y hasta que punto pueden variar algunas métricas en algunos casos como puede ser el split mínimo, aun siendo criterios similares en redes sencillas.

Capítulo 7

Conclusiones y líneas futuras

Como se ha explicado anteriormente, el objetivo de este trabajo era crear una herramienta que nos permitiera conocer de qué forma se pueden distribuir las funciones de red de las distintas estaciones de radio de una red dada, según las condiciones y características de la misma. En el anterior capítulo, hemos podido observar como las distintas tecnologías de enlace, así como los distintos tamaños de red y los criterios según los cuales decidimos la distribución, afectan a esa distribución de las funciones de red.

El análisis realizado ha sido sobre un número bastante pequeño de topologías de red distintas, por lo que los resultados de este trabajo sirven más que nada para hacernos a la idea de cuales pueden ser las tendencias, pero la herramienta desarrollada permitiría ampliar el análisis tanto como fuera necesario, para poder comparar los resultados obtenidos al ejecutarla sobre un número mucho mayor de redes, o empleando distintas tecnologías de enlace. Así como analizar su funcionamiento sobre redes existentes en el mundo real.

Por ello, este podría ser uno de los caminos a seguir a partir de aquí, ampliando el análisis para, de este modo, hacernos a la idea de forma mucho más precisa de cuál podría ser la variación real de los resultados, de encontrarnos ante una cantidad de escenarios a analizar mucho mayor, o una mayor variabilidad en las condiciones de estos escenarios.

Por otro lado, al haber creado una herramienta que permite analizar cualquier red deseada, y que selecciona el mejor resultado entre todos los posibles según los criterios seleccionados en cada caso, la otra línea que puede partir de este trabajo es la creación de una alguna herramienta de funcionamiento a tiempo real, que busque resultados estimados empleando algún tipo de algoritmo aproximativo, pudiendo emplear esta para determinar hasta que nivel los resultados de la herramienta a tiempo real se aproximan al mejor resultado que nos devuelve nuestra herramienta.

Bibliografía

- [1] Aleksandra Checko, Henrik L. Christiansen, Ying Yan, Lara Scolari, Georgios Kardaras, Michael S. Berger, and Lars Dittmann (2015) *Cloud RAN for Mobile Networks — A Technology Overview*
- [2] Dujong Lee, Kyusang Lee, Susang Yoo, and June-Koo Kevin Rhee, Member, IEEE (2011) *Efficient Ethernet Ring Mesh Network Design*
- [3] Andres Garcia-Saavedra , Josep Xavier Salvat, Xi Li, and Xavier Costa-Pere (2018) *WizHaul: On the Centralization Degree of Cloud RAN Next Generation Fronthaul*
- [4] Isaac Alegre (2016) *FrontHaul and BackHaul movil*

Apéndice A

Terminos utilizados

En este anexo voy a explicar los términos empleados para explicar las distintas partes del trabajo, y de las redes en sí, de forma que en el caso de que alguno de los conceptos utilizados durante el desarrollo de este trabajo no quede del todo claro, se pueda recurrir a este índice para tener una explicación más precisa del significado y el sentido de las partes mencionadas. El objetivo de esta parte es, por tanto, mantener una terminología constante a lo largo de la explicación del trabajo y el programa, permitiendo que se entienda con la mayor facilidad posible.

En esta sección empezaré explicando primero los términos más amplios, y pasando posteriormente a las partes más pequeñas de la herramienta, de forma que todo tenga sentido dentro de su contexto, y entendiendo a qué parte pertenece cada cosa.

A.1. Herramienta de análisis

Llamaré Herramienta de análisis a todo el código escrito, es decir, tanto a los distintos algoritmos que permiten encontrar una solución sobre una red ya configurada y a la parte encargada de generar la red según las condiciones especificadas en cada caso, como al propio escenario, encargado de controlar el orden de ejecución de estas acciones, y asignar las características con las que se va a configurar la red en cada ejecución del programa.

Es decir, cuando digo que ejecuto la Herramienta de análisis 3 veces, una sobre cada una de las topologías, cada una de estas tres ejecuciones de la herramienta de análisis ejecuta a su vez el programa un total de 3000 veces, generando 3000 redes distintas sobre la misma topología (ya sea cambiando la tecnología de enlace, la asignación de los nodos, o la capacidad y retardo de los enlaces en una misma tecnología y asignación), y ejecutando los algoritmos 3000 veces (12000 en el caso del encargado de encontrar la solución final, al ser ejecutado una vez por cada uno de los criterios), y encontrando todas las soluciones para esa topología en concreto.

A.2. Programa

Llamamos programa a la suma de la acción encargada de generar una red concreta sobre una topología y la encargada de analizar la red generada, es decir, la acción que controla la ejecución de los algoritmos que obtienen 4 soluciones sobre la red generada. En este caso, cuando decimos que se ejecuta este programa 3000 veces por cada topología, nos referimos a que

se generan 3000 redes distintas y se encuentran las mejores soluciones según los distintos criterios de búsqueda para cada una de ellas.

A.3. Escenario

El escenario es la parte principal de la herramienta, pues es el encargado de controlar la ejecución del resto de sus partes, según los datos que extrae de un fichero de configuración. El escenario incluye dos ciclos anidados que van variando la tecnología de enlace y la asignación de los nodos a lo largo de sus ejecuciones, y dentro de los cuales se ejecuta primero la acción encargada de la creación de la red y luego el análisis de esta sobre la topología introducida en la herramienta.

A.4. Creación de la red

Cada vez que el escenario ejecuta el programa, la primera acción a la que llama es a la creación de la red, la cuál a su vez se divide en la asignación de los nodos y de las capacidades de los enlaces. De esta forma, el programa genera una red sobre la topología introducida en la herramienta de análisis, para posteriormente analizarla en la segunda acción que conforma el programa.

A.5. Asignación de los nodos

La asignación de los nodos es la primera parte del programa. Cada vez que la herramienta de análisis ejecuta el programa, lo primero que hace es generar una fotografía a analizar (es decir, una red sobre la topología de red a introducida en la herramienta), con la tecnología de enlace que toque según el ciclo en el que nos encontremos. Pero antes de generar los enlaces de la red a analizar, la acción encargada de generar la red asigna de que tipo es cada uno de los nodos de forma aleatoria en función del número de nodos de cada tipo indicados en el fichero de configuración.

Es decir, si por ejemplo en la topología 3 tenemos 11 nodos, de los cuales 3 son estaciones de radio, y 2 nodos en la nube, esta parte de la acción podría decidir que los nodos 2, 5 y 7 sean estaciones de radio, el 1 y 4 estaciones en la nube, y el resto nodos intermedios. De esta forma, diremos que cada una de estas configuraciones es una asignación de los nodos.

Uno de los ciclos que encontramos dentro del escenario provoca que para cada topología de red sobre la que se ejecuta la herramienta se realicen 10 asignaciones de los nodos, pues ese ciclo va variando la semilla que se pasa a la acción encargada de generar la red, a partir de la cuál determina de forma pseudo-aleatoria la asignación de estos nodos.

A.6. Algoritmos

Una vez generada la red a analizar (o fotografía), es decir, cuando ya se han asignado los tipos de nodos y las capacidades de los enlaces, la parte del programa realmente encargada de analizar esa red y encontrar la solución, son precisamente los algoritmos. Por tanto, si decimos que estamos analizando una red, estaremos haciendo referencia a la ejecución de los distintos algoritmos sobre una red concreta ya generada, de la cual se van a obtener 4 soluciones, una por cada uno de los 4 criterios empleados.

A.7. Criterios

Cuando mencionamos los criterios, nos referimos a los criterios de selección, es decir, las distintas características y métricas que comparamos entre cada una de las posibles soluciones del árbol, a la hora de seleccionar la mejor de estas.

Como ya se ha explicado durante el trabajo, se han empleado 4 criterios de selección en función de distintas métricas de delay y nivel de centralización, y cuando hablamos de analizar las soluciones para cada uno de los criterios, obviamente nos referimos a comparar el conjunto de todas estas soluciones, que generalmente se corresponderán con los 3000 análisis resultado de multiplicar las 10 asignaciones de los nodos, por las 3 tecnologías de enlace, por las 100 ejecuciones realizadas por cada una de estas combinaciones de asignación y tecnología de enlace asignando de forma aleatoria las capacidades de los enlaces en función de su tecnología.

A.8. Árbol de soluciones

El árbol de soluciones es la forma de representar y guardar dentro del programa todas las soluciones válidas para una red que estamos analizando. El mapa se guarda dentro de nuestro programa como un vector de soluciones, siendo estas soluciones un nuevo tipo de elemento o estructura que guarda una serie de datos correspondientes a cada una de estas soluciones. Para entender más claramente como se conforma este árbol de soluciones, hay que explicar previamente varios conceptos más básicos, como son que es un vector, un mapa, una estructura, una solución o una decisión.

A.8.1. Vector

Un vector es un conjunto ordenado de elementos. En el momento en el que se declara el vector, es necesario indicar el tipo del que serán estos elementos, para posteriormente poder introducirlos en el mismo. En el caso del árbol de soluciones, la forma de guardar sus datos en el programa, es mediante un vector de soluciones. En este caso, una solución es un tipo de elemento creado por nosotros con el tipo de datos concretos que necesitamos para que funcione el algoritmo que encuentra la mejor solución, y poder analizar las características de estas distintas soluciones.

A.8.2. Mapa

Un mapa, al igual que un vector, es un contenedor de elementos. La diferencia principal entre un vector y un mapa, es que este tiene una clave que hace referencia a cada uno de los elementos que contiene. Generalmente una letra, un número o un nombre, con el cual además de poder ordenar los elementos, podemos buscar un elemento concreto dentro del mapa utilizando su clave.

Mientras que los vectores se emplean en casos en los que podemos conocer cual es cada uno de los elementos por su orden, o en el que no nos importa cuál sea cada uno de los elementos, pues lo que nos importa son sus datos, y no a qué elemento se corresponda. Los mapas se utilizan cuando queremos poder identificar a qué o a quién corresponde cada uno de los datos.

Por ejemplo, en este caso, el árbol de soluciones es un vector, pues no le hemos dado un nombre a cada una de las soluciones, y lo que realmente nos importa son las características de estas a la hora de escoger la mejor, mientras que dentro de cada una de las soluciones, encontramos un mapa que guarda la ocupación de los distintos enlaces de la red provocada por la misma, y en ese caso,

necesitamos conocer cuál es el nombre o referencia de los enlaces que han sido ocupados, pues nos interesa saber sobre qué partes de la red se provoca esa ocupación.

A.8.3. Estructura

A la hora de guardar y emplear elementos en nuestro programa, no siempre nos conviene emplear los elementos generales como strings o ints, sino que dependiendo de la situación, podemos necesitar emplear otro tipo de elementos, que a su vez nos permitan guardar varios parámetros concretos según el caso. En esta situación es donde las estructuras cobran sentido, pues permiten crear nuestros nuevos tipos de elementos, de forma que podamos generar elementos que a su vez contengan los tipos de datos que necesitemos en cada caso.

En esta parte del programa, hemos empleado dos tipos de estructuras llamados solución y decisión, que como se puede entender, se corresponden a una solución y una decisión del árbol, y por tanto, nos permiten guardar los datos correspondientes a estas.

A.8.4. Solución

Una solución es el conjunto de las decisiones tomadas para cada una de las estaciones de radio de la red a analizar. Es decir, una de las formas válidas de asignar los caminos y niveles de split a todas las estaciones de radio de la red analizada.

En el programa, una solución es un nuevo tipo de elemento, definido por una estructura que a su vez contiene otros dos elementos, que son un vector que guarda todas las decisiones para cada una de las estaciones de radio, y un mapa que guarda la ocupación de cada uno de los enlaces de la red provocada por todas las decisiones que encontramos en esta solución.

En este caso, se emplea un vector para las decisiones, puesto que estas se van asignando a las distintas estaciones de radio siempre en el mismo orden (RU1, RU2, RU3...), mientras que en el mapa de enlaces, guardamos el nombre de cada uno de estos (es decir, su origen y destino), junto con su ocupación, para poder saber qué enlaces son los que se emplean para enviar los datos.

A.8.5. Decisión

Una decisión es el conjunto de un nivel de split, y un camino para una estación de radio de la red. El conjunto de decisiones válidas para cada una de las estaciones de radio de la red que estamos analizando, conforman una solución, o dicho de otra forma, una solución está formada por una decisión para cada una de las estaciones de radio de la red.

En el programa, estas decisiones son también un nuevo tipo de elemento, que guarda el nivel de split asignado a esa estación de radio, y el delay del camino escogido para realizar esa centralización, pues son esos los datos que nos serán relevantes a la hora de calcular las métricas de cada solución sumando cada una de las decisiones de la misma, teniendo en cuenta además, que el estado de la red se guarda dentro de la propia solución.

A.8.6. Explicación del árbol de soluciones

Una vez desgranados, y explicados los distintos elementos que conforman el árbol de soluciones, es más fácil explicar el mismo. Como ya he dicho antes, en el programa, el árbol de soluciones está guardado como un vector de soluciones, es decir, que cada rama, desde el principio hasta

el final, será guardada como un elemento de ese vector, o una solución, y a la hora de analizar estas ramas y de comparar sus datos en el algoritmo encargado de encontrar la mejor solución, el programa realmente no recorre el árbol decisión a decisión avanzando y retrocediendo por las distintas ramas, sino que simplemente accede a uno elemento de ese vector, lee los datos de la solución correspondiente, y los compara con la mejor solución actual.

Eso sí, en la estructura del elemento solución, podemos observar que no se guardan directamente las métricas a analizar como uno de los elementos de esa estructura, sino que tenemos que acceder al vector de decisiones, para a partir de ahí, sumando los datos de todas las decisiones que conforman nuestra solución, extraer los valores que queremos comparar.

Por tanto, el árbol de soluciones, es un vector de soluciones, cada una de las cuales está formada por un vector de decisiones, y por un mapa que guarda la ocupación de los enlaces. Y a su vez, cada una de las decisiones de este vector, contiene el delay y el nivel de centralización para cada una de las estaciones de radio de la red, y será es a partir de los datos de este vector de decisiones, de dónde se pueden extraer las métricas de cada solución, que posteriormente serán comparados entre si.

A.9. Métricas

Las métricas como concepto son los datos que nos permiten analizar el funcionamiento de un programa u herramienta. En este caso, cuando hablamos de métricas, o métricas de una solución, hacemos referencia a distintos valores correspondientes a esta, que vamos a emplear tanto para escoger la mejor solución para una red concreta, como para posteriormente, comparar las distintas soluciones obtenidas según la topología de red, la tecnología de enlace, y el criterio de selección escogidos en cada caso.

A.10. Ejecución

En general, cuando hablamos de ejecutar, nos referimos a realizar todos los pasos necesarios para completar una acción. Sin embargo, cuando hablamos de ejecuciones, nos podemos referir a la herramienta de análisis al completo, a una ejecución del programa, a una acción de esta, o a un único algoritmo, y por tanto, es necesario entender y aclarar de qué estamos hablando en cada caso.

Cuando hablamos de una ejecución de la herramienta de análisis, nos referimos a la ejecución la herramienta completa para una topología de la red. En este, caso, como ya se ha explicado, durante este análisis, únicamente se ha ejecutado la herramienta de análisis tres veces, una por cada una de las tres topologías de red a analizar, y cada una de estas incluye a su vez, 3000 ejecuciones del programa.

Con esto, lo siguiente es explicar una ejecución del programa, que como se dijo antes, son todos los pasos desde la creación de una red concreta sobre la topología indicada, hasta la obtención de las 4 soluciones para esta.

Por último, cuando hablamos de la ejecución de una acción o un algoritmo, nos referimos a la ejecución exclusiva de esa parte del programa. Por ejemplo, cuando el programa ejecuta el algoritmo Paths, realiza todos los pasos dentro de ese algoritmo para encontrar todos los caminos para las distintas estaciones de radio de la red.

A.11. Acción

El programa, que se ejecuta 3000 veces por cada ejecución de la herramienta de análisis, está formado a su vez por dos acciones, que el escenario se encarga de ejecutar en orden cada vez que ejecuta el programa. Estas acciones son un conjunto de algoritmos y funciones, que permiten realizar los distintos pasos para llegar a los resultados.

Concretamente, la primera acción es la encargada de generar la red, asignando los nodos, y dando capacidades y delays a los enlaces, según toque en cada ejecución del programa. Y la segunda, se encarga de ejecutar los distintos algoritmos que sobre esta red buscan los caminos desde las distintas estaciones de radio a la nube, generan el árbol de soluciones, y en resumen, seleccionan las mejores solución para la red generada por la anterior acción. Como se puede entender, por cada una de las ejecuciones del programa, el escenario se encarga de ejecutar estas dos acciones en orden, generando una nueva red, y encontrando las mejores soluciones para esta en función de los distintos criterios.

A.12. Topología

Una topología es un conjunto de nodos y enlaces que conectan estos nodos entre sí, o explicado de otra forma, es el dibujo sobre el que se genera la red asignando de qué tipo son cada uno de los nodos, y cuales son las características de los distintos enlaces.

En el programa hemos analizado una serie de redes generadas sobre 3 topologías distintas, cada una de las cuales tenía un distinto número de nodos que se encontraban conectados entre sí de una forma determinada. Las topologías empleadas se pueden ver en el anterior capítulo, y como se ha explicado antes, sobre cada una de ellas se generaron un total de 3000 redes distintas a analizar, para extraer 4 soluciones por cada una de estas. Concretamente, en una topología, se asignan los tipos de los nodos de 10 formas distintas, en cada una de las cuales se asignan las características de los enlaces de otras 300 formas, correspondientes a multiplicar las 3 tecnologías de enlace empleadas por 100 asignaciones distintas de sus capacidades por cada asignación de los nodos y tecnología de enlace. De forma que se extraen un total de 12.000 soluciones por topología.

A.13. Red

Una red es un conjunto de nodos y enlaces que permiten la transferencia de datos entre estos nodos. En la ejecución de este programa, las redes representan el Fronthaul de una red de telefonía móvil de quinta generación o 5G, concretamente, representan la parte de la red que conecta las distintas estaciones de radio de una zona concreta con los nodos en la nube, a través de una serie de enlaces y de nodos intermedios.

Por tanto, al emplear el término red en la herramienta, estamos haciendo a una de las redes finales que vamos a analizar, una vez se han asignado los nodos y generado el valor de los enlaces sobre una topología concreta, de forma que el programa se encarga de encontrar las 4 mejores soluciones según los distintos criterios de selección para cada una de estas redes. De esta forma, las soluciones encontradas para las distintas redes, se podrán clasificar en función de la topología a la que se corresponda la red de las que han sido extraídas, y la tecnología de enlace empleada en cada caso.

A.14. Fotografía

En la herramienta, llamamos fotografía a cada imagen a analizar de la que vamos a extraer soluciones. Es decir, que cada vez que se ejecuta el programa, su primera acción genera una fotografía nueva sobre la topología introducida en la herramienta, y posteriormente busca soluciones sobre esta, que posteriormente podrán ser analizadas. Por tanto se puede entender que la red generada, junto con la información de su tecnología de enlace y la topología a la que corresponde, conforma una de estas fotografías.

A.15. Nodos

Como definición general, un nodo es un elemento que sirve de punto de intersección o de nexo entre otros elementos que confluyen en un mismo lugar, lo que en este caso, se entiende como que un nodo es un punto al que se conectan uno o varios enlaces. En el caso de las redes de comunicaciones, los nodos son las distintas máquinas o elementos que se encuentran conectados por enlaces, y permiten y controlan la comunicación y el envío de información a través de estas redes.

En el caso del programa, los nodos son las entidades principales que forman nuestras redes a analizar junto con los enlaces, y tenemos tres tipos de nodos, cada uno de los cuales tienen funciones y objetivos distintos, y que además trataremos de forma distinta durante nuestro análisis, que son:

A.15.1. RUs

Las estaciones de radio o RUs (del inglés Radio Units), son las que permiten que los usuarios se conecten a la red mediante sus dispositivos móviles. En este análisis, el objetivo es analizar cual es la mejor forma de centralizar las funciones de red de cada una de las estaciones de radio de nuestra red, teniendo en cuenta la capacidad y retardo de los enlaces que conectan estas con alguno de los nodos de la nube.

A.15.2. CUs

Los nodos en la nube o CUs (del inglés Cloud Units), son entidades que representan a la nube, y con ello permiten ejecutar las funciones de red centralizadas desde las distintas estaciones de radio de la red, si se cumplen ciertas condiciones de capacidad y retardo en el camino que va desde estas a estos nodos.

A.15.3. INTs

Los nodos intermedios o INTs, son nodos que se encuentran en el camino entre las RUs y las CUs, y permiten el reenvío y el reencaminamiento de los datos a través de los enlaces de la red.

A.16. Funcionamiento

El objetivo del programa, es precisamente, determinar en cuáles de estos nodos de la red se van a realizar las distintas funciones de red. Para esto, analiza las capacidades de los caminos que

conectan las distintas estaciones de radio con alguno de los nodos de la nube, y en función de estos datos, escoge cual es el mejor nivel de centralización (o qué parte de estas funciones se ejecutan en la nube, y cuál en las propias estaciones de radio), para cada una de estas estaciones de radio.

A.17. Enlaces

Un enlace en una red de comunicaciones, es un camino que conecta directamente dos de los nodos de la red. En el caso del programa, empleamos tres tipos distintos de enlaces, cada uno con una capacidad, y una velocidad de propagación distinta, en función de la tecnología empleada, a la hora de generar las distintas redes que vamos a analizar. En cada red generada, eso sí, se empleará únicamente una tecnología de enlace para todos los enlaces de la misma cada vez.

Por otro lado, es importante especificar que aunque he dicho que un enlace es un camino, durante la explicación, cuando emplee el término camino, no me estaré refiriendo a lo mismo que al emplear el término enlace.

A.18. Caminos

Como se acaba de explicar, un camino no es exactamente lo mismo que un enlace. En este caso, mientras que un enlace es un único cable, o una única conexión directa (en el caso de enlaces inalámbricos), entre dos nodos de la red, un camino, es precisamente un conjunto de enlaces que conectan dos nodos de la red. De esta forma, si dos nodos se encuentran conectados directamente, diríamos que el camino que los une tiene un único enlace, pero también pueden estar conectados por varios enlaces, atravesando varios nodos a lo largo de este camino.

En el programa, cuando haga referencia a un camino, generalmente hablaré del camino que conecta una de las estaciones de radio de la red, con uno de los nodos de la nube, de forma que los nodos que hay a lo largo de ese camino, serán generalmente (aunque no en todos los casos), nodos intermedios.

Entendido el concepto de camino, el programa lo que hace es calcular la capacidad y el delay de cada uno de estos caminos, teniendo en cuenta las características de los enlaces que forman cada uno de ellos, y a partir de ahí, analizar los niveles de centralización que se pueden asignar a las distintas estaciones de radio en función tanto de las capacidades de los caminos, como de los enlaces compartidos entre los caminos de distintas RUs.

A.19. Métricas de los enlaces y caminos

A la hora de hablar de los enlaces y los caminos, hay dos métricas de estos que son relevantes, que además son las que determinan el nivel de centralización que se puede asignar a una estación de radio a través de uno de estos caminos. En el caso de los enlaces, estas métricas vendrán indicadas directamente en el momento en el que se genera cada uno de ellos, dentro de los datos del mismo. En cambio, en un camino, las métricas dependerán de todos los enlaces que forman ese camino. Estas métricas son:

A.19.1. Delay

Cuando hablamos del delay o retardo, estamos haciendo referencia al tiempo que cierta información tarda en recorrer un enlace o camino. En el caso de los enlaces, el delay es el resultado de multiplicar la longitud de este, por la velocidad de propagación correspondiente a la tecnología de enlace que se emplee en el camino.

Por otro lado, el delay de un camino, es la suma de los delays de todos los enlaces que forman parte de este camino, una vez hemos configurado el algoritmo encarado de generar los caminos para que no haya bucles ni ciclos en estos, o dicho de otra forma, para que una camino no pase más de una vez por el mismo nodo o enlace.

A.19.2. Capacidad

La capacidad de un enlace o camino, es la cantidad de datos que se pueden enviar por el mismo en un periodo de tiempo. También nos podemos referir a este parámetro como velocidad de transmisión o simplemente velocidad del enlace. En el caso de los enlaces, esta capacidad está determinada por la tecnología de enlace empleada en cada caso, mientras que la capacidad de un camino, será la del enlace con menos capacidad entre todos los que lo forman.

En este caso, al emplear una única tecnología de enlace por cada red analizada, la capacidad inicial de todos los caminos de una red, será siempre la misma. Pero debido a que a la hora de determinar que soluciones, combinaciones de decisiones son válidas para las distintas estaciones de radio de la red, los caminos de estas pueden compartir enlaces, hay que tener en cuenta esta limitación. Por este motivo, podría darse el caso de que si los caminos correspondientes a dos estaciones de radio comparten un enlace, y una de las dos ocupa tres cuartas partes de su capacidad, la capacidad del camino para la segunda, estaría limitada a una cuarta parte de la inicial, al tener un enlace que tiene ocupadas las otras tres cuartas partes.

A.20. Nivel de centralización

El nivel de centralización, también llamado nivel de split, o simplemente split, es precisamente el punto central de este trabajo. Como se ha explicado extensamente en este trabajo, en las redes 5G, las funciones de red pueden ejecutarse tanto en las propias estaciones de radio, como en la nube, y el nivel de centralización es la cantidad de estas funciones que se mandan a la nube para reducir la carga de trabajo de una estación de radio.

Al emplear estos términos, podemos referirnos a la centralización de una estación de radio específica, o del conjunto de las estaciones de radio de la red. Y en cualquier caso, este nivel de centralización, dependerá por encima de todo de las características de los caminos que conectan nuestras estaciones de radio con la nube.